

THE CAPCO INSTITUTE
JOURNAL
OF FINANCIAL TRANSFORMATION



BALANCING
INNOVATION & CONTROL

a **wipro** company

#59 JUNE 2024

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

Editor

Shahin Shojai, Global Head, Capco Institute

Advisory Board

Lance Levy, Strategic Advisor

Owen Jelf, Partner, Capco

Suzanne Muir, Partner, Capco

David Oxenstierna, Partner, Capco

Editorial Board

Franklin Allen, Professor of Finance and Economics and Executive Director of the Brevan Howard Centre, Imperial College London and Professor Emeritus of Finance and Economics, the Wharton School, University of Pennsylvania

Philippe d'Arvisenet, Advisor and former Group Chief Economist, BNP Paribas

Rudi Bogni, former Chief Executive Officer, UBS Private Banking

Bruno Bonati, Former Chairman of the Non-Executive Board, Zuger Kantonalbank, and President, Landis & Gyr Foundation

Dan Breznitz, Munk Chair of Innovation Studies, University of Toronto

Urs Birlcher, Professor Emeritus of Banking, University of Zurich

Elena Carletti, Professor of Finance and Dean for Research, Bocconi University, Non-Executive Director, UniCredit S.p.A.

Lara Cathcart, Associate Professor of Finance, Imperial College Business School

Géry Daeninck, former CEO, Robeco

Jean Dermine, Professor of Banking and Finance, INSEAD

Douglas W. Diamond, Merton H. Miller Distinguished Service Professor of Finance, University of Chicago

Elroy Dimson, Emeritus Professor of Finance, London Business School

Nicholas Economides, Professor of Economics, New York University

Michael Enthoven, Chairman, NL Financial Investments

José Luis Escrivá, President, The Independent Authority for Fiscal Responsibility (AIReF), Spain

George Feiger, Pro-Vice-Chancellor and Executive Dean, Aston Business School

Gregorio de Felice, Head of Research and Chief Economist, Intesa Sanpaolo

Maribel Fernandez, Professor of Computer Science, King's College London

Allen Ferrell, Greenfield Professor of Securities Law, Harvard Law School

Peter Gomber, Full Professor, Chair of e-Finance, Goethe University Frankfurt

Wilfried Hauck, Managing Director, Statera Financial Management GmbH

Pierre Hillion, The de Picciotto Professor of Alternative Investments, INSEAD

Andrei A. Kirilenko, Reader in Finance, Cambridge Judge Business School, University of Cambridge

Katja Langenbacher, Professor of Banking and Corporate Law, House of Finance, Goethe University Frankfurt

Mitchel Lenson, Former Group Chief Information Officer, Deutsche Bank

David T. Llewellyn, Professor Emeritus of Money and Banking, Loughborough University

Eva Lomnicka, Professor of Law, Dickson Poon School of Law, King's College London

Donald A. Marchand, Professor Emeritus of Strategy and Information Management, IMD

Colin Mayer, Peter Moores Professor of Management Studies, Oxford University

Francesca Medda, Professor of Applied Economics and Finance, and Director of UCL Institute of Finance & Technology, University College London

Pierpaolo Montana, Group Chief Risk Officer, Mediobanca

John Taysom, Visiting Professor of Computer Science, UCL

D. Sykes Wilford, W. Frank Hipp Distinguished Chair in Business, The Citadel

CONTENTS

GOVERNANCE OF TECHNOLOGY

- 08 Data and AI governance**
Sarah Gadd, Chief Data Officer, Bank Julius Baer
- 20 “Data entrepreneurs of the world, unite!” How business leaders should react to the emergence of data cooperatives**
José Parra-Moyano, Professor of Digital Strategy, IMD
- 26 Revolutionizing data governance for AI large language models**
Xavier Labrecque St-Vincent, Associate Partner, Capco
Varenya Prasad, Principal Consultant, Capco
- 32 Municipal data engines: Community privacy and homeland security**
Nick Reese, Cofounder and COO, Frontier Foundry Corporation
- 40 Human/AI augmentation: The need to develop a new people-centric function to fully benefit from AI**
Maurizio Marcon, Strategy Lead, Analytics and AI Products, Group Data and Intelligence, UniCredit
- 50 Building FinTech and innovation ecosystems**
Ross P. Buckley, Australian Research Council Laureate Fellow and Scientia Professor, Faculty of Law and Justice, UNSW Sydney
Douglas W. Arner, Kerry Holdings Professor in Law and Associate Director, HKU-Standard Chartered FinTech Academy, University of Hong Kong
Dirk A. Zetzsche, ADA Chair in Financial Law, University of Luxembourg
Lucien J. van Romburg, Postdoctoral Research Fellow, UNSW Sydney
- 56 Use and misuse of interpretability in machine learning**
Brian Clark, Rensselaer Polytechnic Institute
Majeed Simaan, Stevens Institute of Technology
Akhtar Siddique, Office of the Comptroller of the Currency
- 60 Implementing data governance: Insights and strategies from the higher education sector**
Patrick Cernea, Director, Data Strategy and Governance, York University, Canada
Margaret Kierylo, Assistant Vice-President, Institutional Planning and Chief Data Officer, York University, Canada
- 70 AI, business, and international human rights**
Mark Chinen, Professor, Seattle University School of Law

GOVERNANCE OF SUSTAINABILITY

82 Government incentives accelerating the shift to green energy

Ben Meng, Chairman, Asia Pacific, Franklin Templeton

Anne Simpson, Global Head of Sustainability, Franklin Templeton

92 Governance of sustainable finance

Adam William Chalmers, Senior Lecturer (Associate Professor) in Politics and International Relations, University of Edinburgh

Robyn Klingler-Vidra, Reader (Associate Professor) in Entrepreneurship and Sustainability, King's Business School

David Aikman, Professor of Finance and Director of the Qatar Centre for Global Banking and Finance, King's Business School

Karlygash Kuralbayeva, Senior Lecturer in Economics, School of Social Science and Public Policy, King's College London

Timothy Foreman, Research Scholar, International Institute for Applied Systems Analysis (IIASA)

102 The role of institutional investors in ESG: Diverging trends in U.S. and European corporate governance landscapes

Anne LaFarre, Associate Professor in Corporate Law and Corporate Governance, Tilburg Law School

112 How banks respond to climate transition risk

Brunella Bruno, Tenured Researcher, Finance Department and Baffi, Bocconi University

118 How financial sector leadership shapes sustainable finance as a transformative opportunity: The case of the Swiss Stewardship Code

Aurélia Fäh, Senior Sustainability Expert, Asset Management Association Switzerland (AMAS)

GOVERNANCE OF CORPORATES

126 Cycles in private equity markets

Michel Degosciu, CEO, LPX AG

Karl Schmedders, Professor of Finance, IMD

Maximilian Werner, Associate Director and Research Fellow, IMD

134 Higher capital requirements on banks: Are they worth it?

Josef Schroth, Research Advisor, Financial Stability Department, Bank of Canada

140 From pattern recognition to decision-making frameworks: Mental models as a game-changer for preventing fraud

Lamia Irfan, Applied Research Lead, Innovation Design Labs, Capco

148 Global financial order at a crossroads: Do CBDCs lead to Balkanization or harmonization?

Cheng-Yun (CY) Tsang, Associate Professor and Executive Group Member (Industry Partnership), Centre for Commercial Law and Regulatory Studies (CLARS), Monash University Faculty of Law (Monash Law)

Ping-Kuei Chen, Associate Professor, Department of Diplomacy, National Chengchi University

158 Artificial intelligence in financial services

Charles Kerrigan, Partner, CMS

Antonia Bain, Lawyer, CMS



DEAR READER,

In my new role as CEO of Capco, I am very pleased to welcome you to the latest edition of the Capco Journal, titled **Balancing Innovation and Control**.

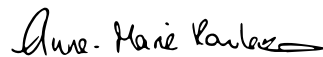
The financial services and energy sectors are poised for another transformative year. At Capco, we recognize that this is a new era where innovation, expertise, adaptability, and speed of execution will be valued as never before.

Success will be determined based on exceptional strategic thinking, and the ability to leverage innovative new technology, including GenAI, while balancing a laser focus on risk and resilience. Leaders across the financial services and energy industries recognize the transformative benefits of strong governance while needing to find the optimal balance between innovation and control.

This edition of the Capco Journal thus examines the critical role of balancing innovation and control in technology, with a particular focus on data, AI, and sustainability, with wider corporate governance considerations. As always, our authors include leading academics, senior financial services executives, and Capco's own subject matter experts.

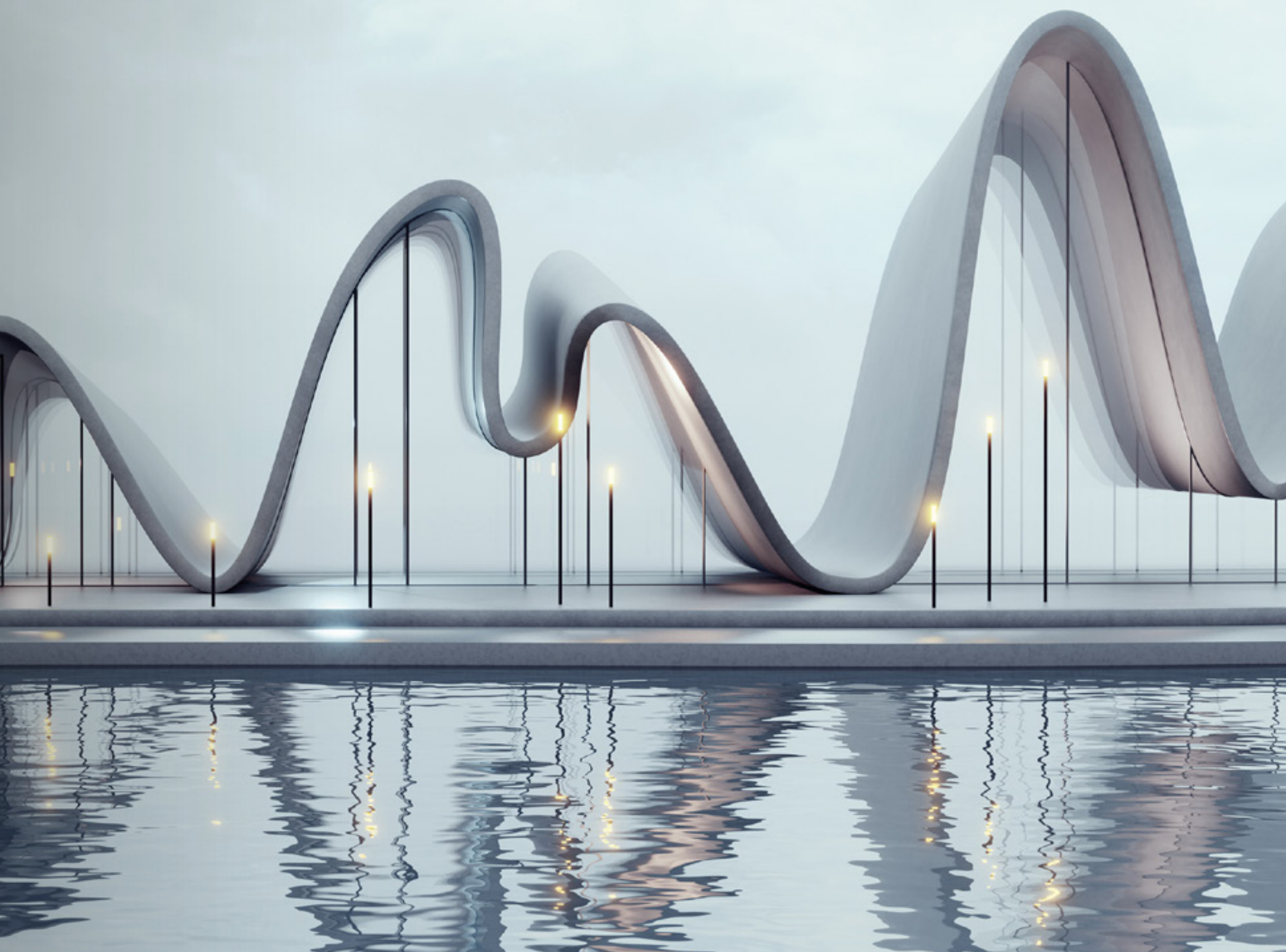
I hope that you will find the articles in this edition truly thought provoking, and that our contributors' insights prove valuable, as you consider your institution's future approach to managing innovation in a controlled environment.

My thanks and appreciation to our contributors and our readers.



Annie Rowland, **Capco CEO**

GOVERNANCE OF TECHNOLOGY



08 Data and AI governance

Sarah Gadd, Chief Data Officer, Bank Julius Baer

20 “Data entrepreneurs of the world, unite!” How business leaders should react to the emergence of data cooperatives

José Parra-Moyano, Professor of Digital Strategy, IMD

26 Revolutionizing data governance for AI large language models

Xavier Labrecque St-Vincent, Associate Partner, Capco

Varenya Prasad, Principal Consultant, Capco

32 Municipal data engines: Community privacy and homeland security

Nick Reese, Cofounder and COO, Frontier Foundry Corporation

40 Human/AI augmentation: The need to develop a new people-centric function to fully benefit from AI

Maurizio Marcon, Strategy Lead, Analytics and AI Products, Group Data and Intelligence, UniCredit

50 Building FinTech and innovation ecosystems

Ross P. Buckley, Australian Research Council Laureate Fellow and Scientia Professor, Faculty of Law and Justice, UNSW Sydney

Douglas W. Arner, Kerry Holdings Professor in Law and Associate Director, HKU-Standard Chartered FinTech Academy, University of Hong Kong

Dirk A. Zetzsche, ADA Chair in Financial Law, University of Luxembourg

Lucien J. van Romburg, Postdoctoral Research Fellow, UNSW Sydney

56 Use and misuse of interpretability in machine learning

Brian Clark, Rensselaer Polytechnic Institute

Majeed Simaan, Stevens Institute of Technology

Akhtar Siddique, Office of the Comptroller of the Currency

60 Implementing data governance: Insights and strategies from the higher education sector

Patrick Cernea, Director, Data Strategy and Governance, York University, Canada

Margaret Kierylo, Assistant Vice-President, Institutional Planning and Chief Data Officer, York University, Canada

70 AI, business, and international human rights

Mark Chinen, Professor, Seattle University School of Law

DATA AND AI GOVERNANCE

SARAH GADD | Chief Data Officer, Bank Julius Baer¹

ABSTRACT

Data governance has come a long way from its inception in the 1980s, transitioning from a necessary overhead to a vital business capability enabling intelligence at scale. This article discusses the data governance journey to data governance 3.0, the role data products can play in risk-managed business self-service with a future view, and the lessons we can learn that will help move AI governance from infancy to value enabler at scale.

1. INTRODUCTION

Data governance – involving Excel spreadsheets and checklists to capture the business concepts represented by the data – has been around since the 1980s. It was viewed then as a “necessary overhead” and had no link back to the actual data. In essence, as Hinkle (2020) notes, it was “a process for cataloging large quantities of transactional data.”

A Chief Data Officer’s role in that foundational period was to simply collate concepts and create inventories of these concepts. Updates were done infrequently, sometimes annually, through manual reviews, while data ownership was seen as a “technology problem” with little in the way of business accountability for the data being created.

This status quo remained in place until the early 2000s, right up to when the “digital transformation” and the “big data frenzy” came into being. This quickly led to what became known as “data governance 2.0” – essentially, to a new paradigm where “data as an asset” principles were created to enable modern, data-driven businesses.

Distilled, this new era can be explained by the phrase coined by Clive Humby in 2006: “Data is the new oil”, which like oil, is “valuable, but if unrefined it cannot really be used” [Watts (2021), Talagala (2022)]. Data governance 2.0 embraced collaboration, broke down organizational silos, and spread accountability across more data governance specific roles alongside business ownership.

In 2018, the Wall Street Journal ran the headline “Global reckoning on data governance” [Loftus (2018)]. That was the time when data breaches at a number of global organizations resulted in decreased revenues due to reputational damage, making headlines around the world. On May 25th, 2018 the E.U.’s General Data Protection Regulation (GDPR) came into effect [E.U. (2018)], leaving many companies struggling to meet compliance standards.

That same year also saw artificial intelligence (AI) governance become a hot regulatory topic, with the European Commission working on developing the “Assessment list for trustworthy AI” (ALTAI), released in June 2020 [E.C. (2020)]. At the end of 2019, the Hong Kong Monetary Authority (HKMA) published a report titled “Reshaping banking with artificial intelligence” [HKMA (2019)], as part of a series of studies on the opportunities and challenges of applying AI technology in the banking industry. The Bank of England and the Financial Conduct Authority launched the “Artificial intelligence public-private forum” (AIPPF) on October 12th, 2020. On April 21st, 2021, the AI Act was officially proposed, with an agreement being concluded on December 9th, 2023 [European Parliament (2023)], while the Monetary Authority of Singapore published a toolkit for assessment of AI by financial institutions in June 2023 [MAS (2023)].

¹ Contributor: Bea Schroettner, Certified Data Ethicist, Bank Julius Baer. Edited by: Natalie Martini.

The result? Nations across the world are either updating existing regulations on data privacy and copyright, looking to create new AI specific regulations, or are searching for ways to embrace guiding principles such as the G7 AI Code of Conduct (currently in development) [OECD (2023)].

We have now entered the era of “data governance 3.0”. What does this look like?

At its core, this is about utilizing data science and improved technologies to treat data governance as a true enabler for organizations. Large language models (LLMs), AI, and active metadata,² breathe life into all of the artifacts that were captured over the last two decades. Data governance 3.0 is a living part of the organization, improving efficiency through integration and automation. Compliance, data quality, and effective data management are built in by design, not add-ons at the end of a process.

But what is “AI governance 1.0”?

In essence, this is about building the foundations that will enable safe, ethical, scalable use of AI, in a world of fast-evolving regulation and technology.

Exponentially increasing unstructured data volumes, computing power, and citizen analytics and data science capabilities, offer organizations the treasure of more and more data intelligence. But this all comes at a cost. As we saw in 2018, when data governance faced a global reckoning, the risks associated with providing AI tools without the culture or the knowledge is elevated. The hard lessons that were learnt from the data governance journey need to be implemented if we are to evolve AI governance. Focus needs to be on education, culture, and strategic alignment as key facets of successful AI governance.

In short, it is not just about governing the model underlying the AI solution. AI governance is everyone’s role. Governance must operate in the delicate balance between regulation and risk mitigation on one side and enablement and innovation on the other.

If this balance is achieved, well-designed governance can generate tangible value while evolving with a future that remains unknown.

Peter Drucker, one of the 20th century’s leading management theorists, put it well: “The greatest danger in times of turbulence is not the turbulence; it is to act with yesterday’s logic” [McConnell (2020)].

2. DATA GOVERNANCE 3.0

The International Data Management Organization noted: “Data governance is defined as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets. [...] Data governance focuses on how decisions are made about data and how people and processes are expected to behave in relation to data” [DAMA International (2017)].

Implied in this definition is the alignment with a more traditional governance model, which lacks the dimension of what governance should be actively promoting: the desired outcome. In other words, to ensure that discoverable, curated, high-quality data is securely available to users – as and when they need it. Put differently, an “enabler” that brings together high-quality data and consumers of data to deliver trustworthy data-driven insights.

With the rise of big data alongside advances in computing power, the interest in generating insights from data has skyrocketed in the last decade. With the increased importance of data science and data-lead decision making, a range of data topics were pushed into focus, data quality being the most prominent [Brous et al. (2020)]. The fact that data scientists spent, and arguably still spend, a significant amount of their time cleaning and organizing (poorly governed) data [McKinsey (2020)] before any value generation, further highlighted the need to change the data governance approach. At the same time, highly publicized data breaches and failures reiterated in parallel the need for the gatekeeping aspect of the data governance role to become more prominent [Famularo (2019)].

Data governance 3.0 strives to achieve an effective way of balancing risk control with user-enabling innovation and insight generation. The ability to extract high-quality insights from data is maturing from being a competitive advantage to a necessary hygiene factor. George Fuechsel, an IBM programmer and instructor, is generally credited with coining the term “garbage in, garbage out” (GIGO) in the early 1960s [Awati (2023)], and 60 years later it still remains one of main hurdles for enabling data value generation, for both business intelligence as well as generative AI (GenAI).

² Metadata is a set of data that describes and gives information about other data, e.g., whether a piece of data is a personal identifier.

How can we realize the data governance 3.0 benefits? I believe that we need to stop thinking about data as just an “asset” and start thinking of data as a product ingredient, and as with all product ingredients, apply consumer safety standards. The core meaning of data hygiene has not changed, it is still the absolute need to understand the quality of a piece of data and what that piece of data can be used for. What has changed is the ability to use machine learning and LLMs to vastly improve data quality detection alongside robust data classification. One of the high barriers to data insights has been the ability to access the data itself, with estimates of between 50 to 70% of time being spent just getting access to the data you need to answer a question. Data access automation and attribute-based access control can now be realized by converting internal policies into sets of machine-readable rules, which, when overlaid with the attributes of the data consumer, their patterns of data usage, and the attributes of the data, can streamline data access greatly, thus reducing the time to answer the question (i.e., time to insights).

Data governance 2.0 moved from “concepts and cataloging” to physical data, while data governance 3.0 activates the physical data level by using data science approaches to understand data securely at scale. The governance roles that ensure the ownership and accountability for data need to remain in place but demand empowering through technological advancements, not manual exercises. Data governance 3.0 should embrace the use of technology from the moment data is created, to when that data is deleted (the data lifecycle). You need to augment governance through embedding AI/machine learning algorithms in the data lifecycle, so they can do what they are good at: dealing with vast amounts of data to classify, qualify, and enhance. By doing this, you provide the data governance “human in the loop” with fast insights they can use to make informed data governance decisions.

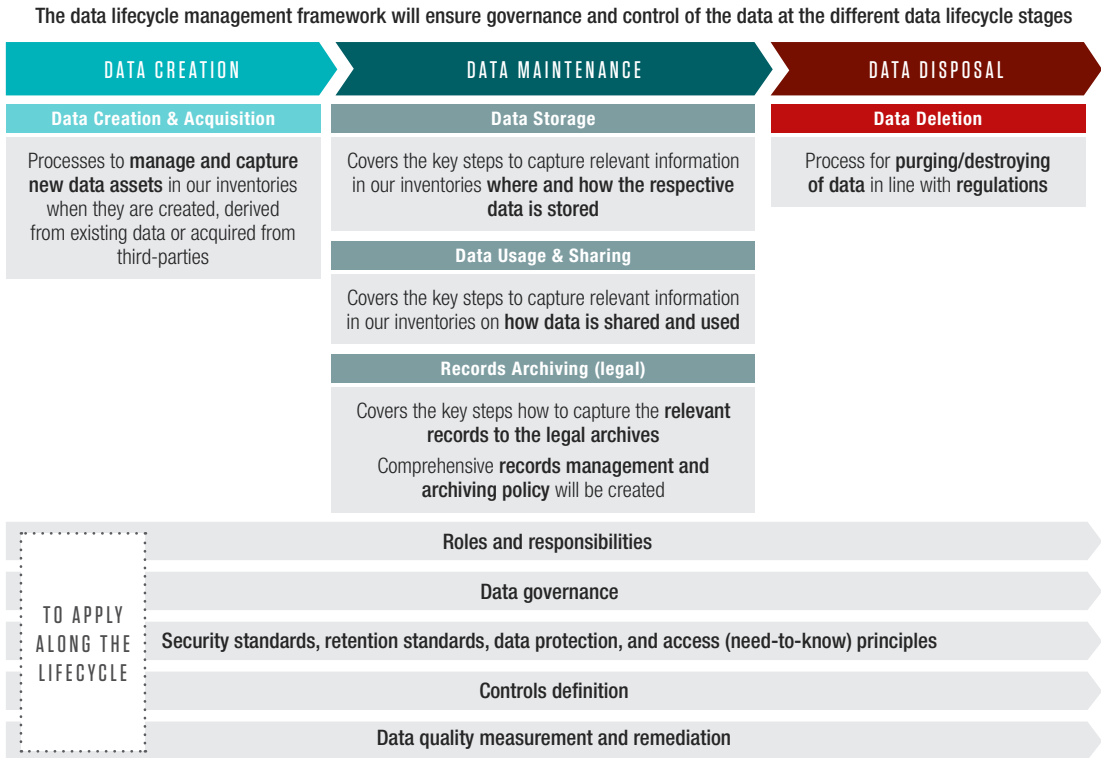
One practical example that applies in many companies is “entity resolution”, i.e., when data is coming in at scale with different identifiers (including names) that actually represent the same thing – like client names, third-party names, and inventory items. There are proven machine learning techniques that will mine the data as it flows in and create clusters with a probability score on the data the machine believes represents the same entity. These clusters are presented to the human to validate and the machine continues to learn. In data governance 2.0, this was an extremely time consuming and often impossible task.

Some aspirational approaches to be considered across the data management lifecycle are (Figure 1):

- **Data creation:** pattern recognition to automatically classify whether a data attribute is a birthdate, social security number, third-party name, client ID number, or other personal identifying information that is deemed critical and needs a higher level of protection.
- **Data storage and archiving:** auto-classification of data (and records) that need to be stored in an archive for regulatory/business purposes and matching them to the length of retention that applies.
- **Data lineage:** tracking of data lineage to identify data that is not from authoritative sources versus the data that is, and where the data is being manipulated so it no longer represents the “truth”.
- **Data usage:** use of a combination of machine-readable controls and attributes of the person trying to access the data (e.g., role, point of time location, normal access patterns, etc.) to provide the data as readable or obfuscated (with patterns intact for data scientists) in the environment needed by the user, whether for development, business intelligence, analytics, or data science.
- **Data quality:** applying AI to help facilitate the improvement of data quality, e.g., through data standardization, data validation, or data governance compliance checks and other features [Drenik (2023)]. Virtually all major data quality management tools already contain this functionality, which provides the standards and guardrails within which domains should operate. Setting up these tools is a critical enablement opportunity for a data governance function.
- **Data deletion:** machine-readable retention rules cross-referenced with legal hold information to enable the compliant on-time deletion of data and records, either from legal archives or from operational systems through API calls.

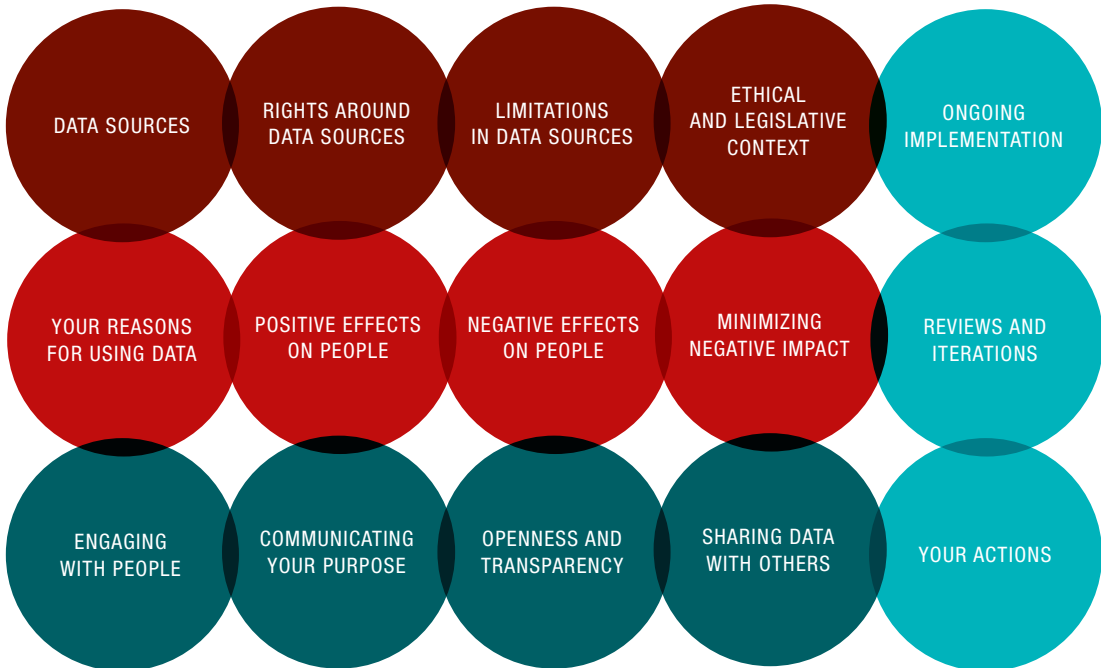
While data governance and data management efforts have traditionally been focused mainly on structured data, expanding this effort to unstructured data (e.g., documents, emails, and contracts) is of growing relevance. The amount of unstructured data is rapidly increasing – some estimate as much as 90% of a company’s data to be unstructured [Violino (2023)]. By its definition, unstructured data does not follow a clear schema or data model and it may contain personal or sensitive data that is harder to spot.

Figure 1: Data lifecycle

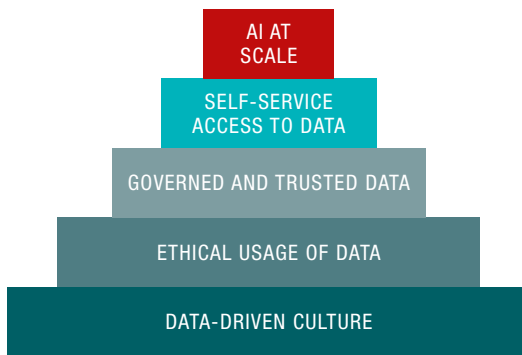


Source: Bank Julius Baer

Figure 2: Data ethics canvas



Source: Derived from Open Data Institute (ODI)
License: CC-BY-SA 4.0

Figure 3: Enabling AI at scale

Source: Bank Julius Baer

GenAI and the power of LLMs have commoditized the capability of extracting value and insights from semi-structured and unstructured data, though trust in outcomes remains a challenge. “Garbage in, garbage out” also applies in the unstructured world, as do the normal dimensions of good data governance, including overall data hygiene [Abdullahi (2023), Rosencrance (2024)]. For example, unstructured data could be of questionable quality (e.g., multiple or duplicate versions of the same document). LLMs have enabled us to better identify and classify the data ingredients within the unstructured world. For instance, you can use models to identify and classify clauses in contracts, sensitive or personal identifying data, start dates, parties, and terms. These features make it far easier to mine data securely for intelligence. One could argue that wrangling unstructured data and applying governance is the larger value proposition and more likely to be a differentiator than the structured world of data.

With the incredible volume of data that enterprises are managing daily, the only way to curate the data ingredients is by using the data science tools and techniques that are available today and constantly evolving with the technology to bring further future value.

While AI ethics is a global topic, it is not a new concept for data. Hasselbalch and Tranberg (2016) was one of the early books to describe not only the privacy implications of the commercial exploitation of big data, but also the broader social and ethical implications. The Open Data Institute published a “data ethics canvas” in 2021, covering many of the aspects that are now in the news with AI ethics (Figure 2) [ODI (2021)].

Culture and ethics are vital aspects for successful data and AI governance and should be treated as critical governance dimensions (Figure 3). Everyone needs basic data literacy and awareness to understand the questions that should be asked when consuming or working with data.

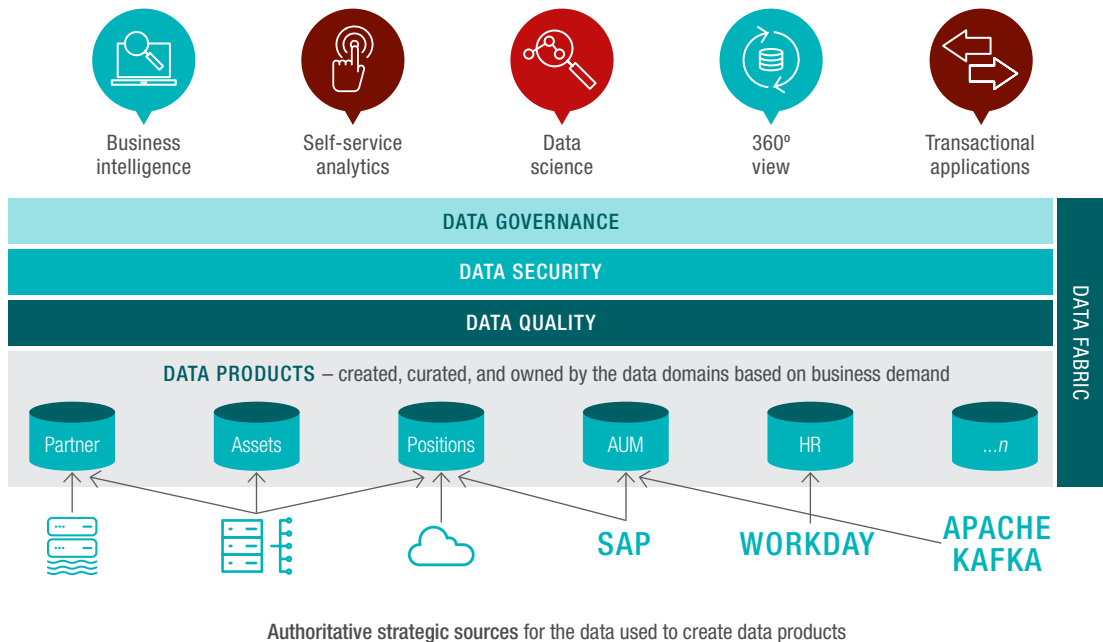
So far, we have covered data governance 3.0 and adopting AI capabilities to build data governance by design across the data management lifecycle, as well as the importance of culture and ethics. Next, we will look at what happens if you do not have the resources and funding to invest in curating data ingredients, or perhaps even if you do. Enter “data products”, which build on data governance 2.0 and 3.0, and in the immediate future will leverage the power of LLMs and GenAI to enable the business to self-serve automated data product creation.

3. DATA PRODUCTS

We have lived through the era of master data management, data warehouses, data lakes, and data lakehouses, and one challenge that consistently arises is “how do I keep all this data in sync”? Data is typically created in places that are fit for that type of data, whether that is software as a service, operational data stores, data integration layers, or even mainframes. The approach of a “one stop place for all data” has not worked, with many enterprises trying and failing [Woods (2016)].

Not all data is equal, and not all data has business value. Many enterprises focus on “critical” or “material” data, which at its core sounds good, but quite often the importance of the data is driven by the need for that data at any given time, which changes based on circumstances at that time.

“Data as a product” (DaaP) first appeared in 2019 as part of the “data mesh” concept defined by Zhamak Dehghani [Fowler (2019)]. Simply, a data product is a broad definition that includes any product or feature that utilizes data to facilitate a goal. Essentially, in addition to (or instead of) using the more manual data governance 2.0 approach, you could apply all the data governance approaches discussed in data governance 3.0 to a grouping of data ingredients rather than each individual ingredient. For example, if I wanted to create a data product that represented sales in the U.S., as the data owner for sales data, I could point to the individual sources for that data (whatever those might be) and put the quality measurement, security, data classification, on the product level instead of the individual attributes, and manage the data collection as a data product.

Figure 4: Example illustration of a data product framework

Source: Bank Julius Baer

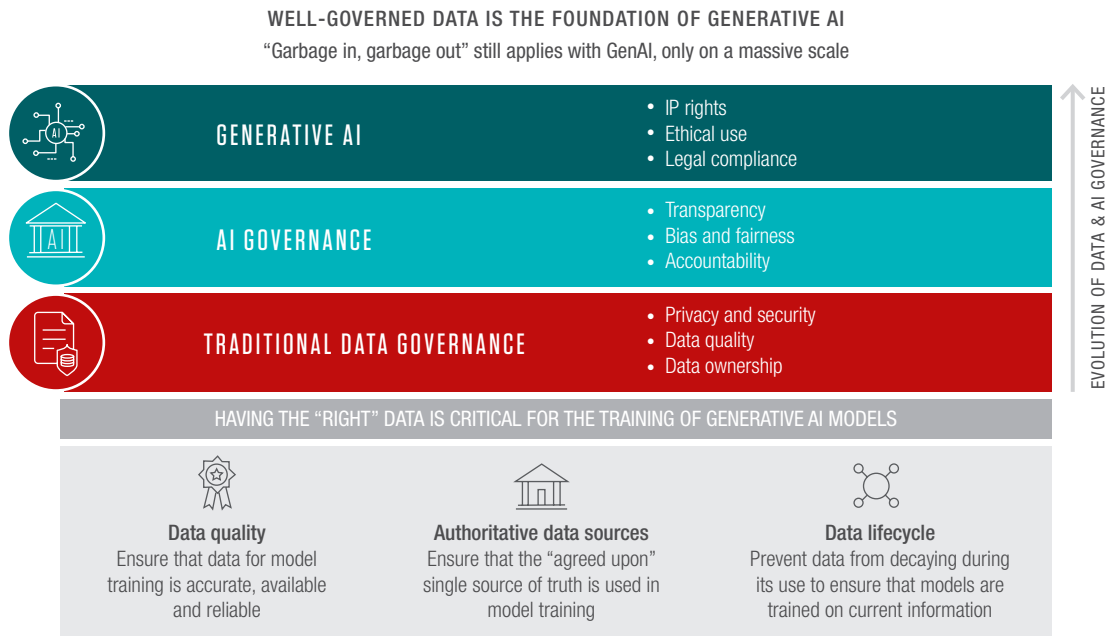
McKinsey & Co. define a data product as delivering “a high-quality, ready-to-use set of data that people across an organization can easily access and apply to different business challenges” [McKinsey (2022)]. Data products offer a number of benefits. First and foremost, they are built, curated, and maintained by subject matter experts (usually the data owners). As such, they represent an “official” version of the data. This reduces the effort for data analysts and data scientists, who would otherwise have to find the right data fields in raw data (data ingredients) and may integrate or manipulate them in a way that provides an answer that is not representative of the question.

Data products also save time through reuse – everyone can securely “shop for the product”, not having to build it for themselves. Data products increase consistency across business intelligence and reporting, as people start with a “common view” and can then combine data products into a new data product to answer another business question. Like other products, data products can be advertised in a product store or a data catalogue and have an owner who can monitor usage and curates them throughout their lifecycle. Well-managed data products can reduce the time to implement use-cases significantly, by up to 90% [Desai et al. (2022)], and strengthen the concept of user self-service.

You can create data products off any kind of data store or combination thereof. This means you do not need to move all your enterprise data to one giant store somewhere, but rather you source the data ingredients for the data product from their “home store”, for example, HR data from systems like Workday, finance data from SAP, etc. (Figure 4). Getting the data from the authoritative store for that data means that you do not need to constantly try to keep copied data in sync with the authoritative store, which, if not done properly, can result in data breaks and stale data being consumed by end-users.

Data governance still plays a major role in the creation of data products by, for example, providing clear standards around “metadata documentation, data classification, and data quality monitoring” [McKinsey (2023a)], as well as ways to enforce governance across the product lifecycle [Deighton (2023)] – where data governance 3.0 can play a role. In addition, the data owner needs to demonstrate data health for their product (just like health standards when you buy food products), otherwise how can users trust the data product they have been given?

The benefits derived from data products not only accrue to the consumers of the product, while driving business value, but also to data governance (Figure 5). They also help reduce risk [McKinsey (2022)]. By controlling how consumers can access

Figure 5: Data, AI, and GenAI governance – how they all fit together

Source: Bank Julius Baer

your products, you can put the right safety standards on the products, which reduces the risk management complexity of doing this across all the data ingredients. This is analogous to shopping in a supermarket: when you buy a tin of soup, you trust the manufacturer, you trust the container, you do not need to review every single ingredient (though they are listed so you have the option), instead you trust that the tin of soup is going to be exactly what you thought it would be.

With the power of GenAI and LLMs to mine metadata and generate code, the ability for the business to create data products as code using natural language is emerging, further commoditizing data product generation with the business owners of the data being able to self-serve. Applying LLMs to a) allow domain experts who have business, but not necessarily coding skills, to specify the data products they wish to build and b) extract the relevant metadata to build the required data pipelines, overlaying security and governance, further builds upon the concepts of data governance by design.

4. AI GOVERNANCE: MACHINE LEARNING TO GenAI

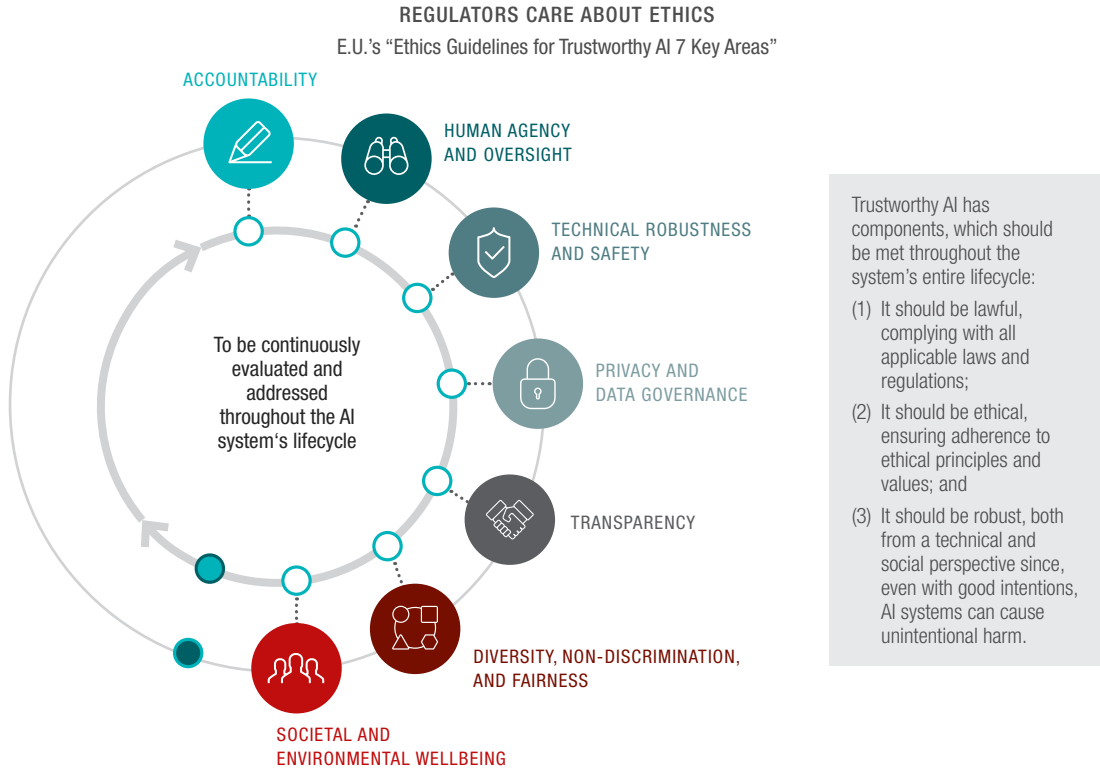
AI governance can be defined as “a system of rules, practices, processes, and technological tools that are employed to ensure an organization’s use of AI technologies aligns with the organization’s strategies, objectives, and values; fulfills legal

requirements; and meets principles of ethical AI followed by the organization” [Birkstedt et al. (2023)]. Part of AI governance is the layer resting on top of data governance, as AI solutions, at their core, consist of input data, the models or algorithms trained for specific tasks, and their output [IBM (n.d.)]. Model input and output are data and as such, benefit from a strong and effective data governance across both structured and unstructured data.

However, AI requires additional facets of governance. AI models often automate decisions and/or processes – due to the increasing complexity of AI solutions, understanding and explaining how a decision was arrived at can be challenging. Model output can sometimes display unwanted bias or could be discriminatory against certain groups. With GenAI, intellectual property violations have been widely reported in the media [Appel et al. (2023)].

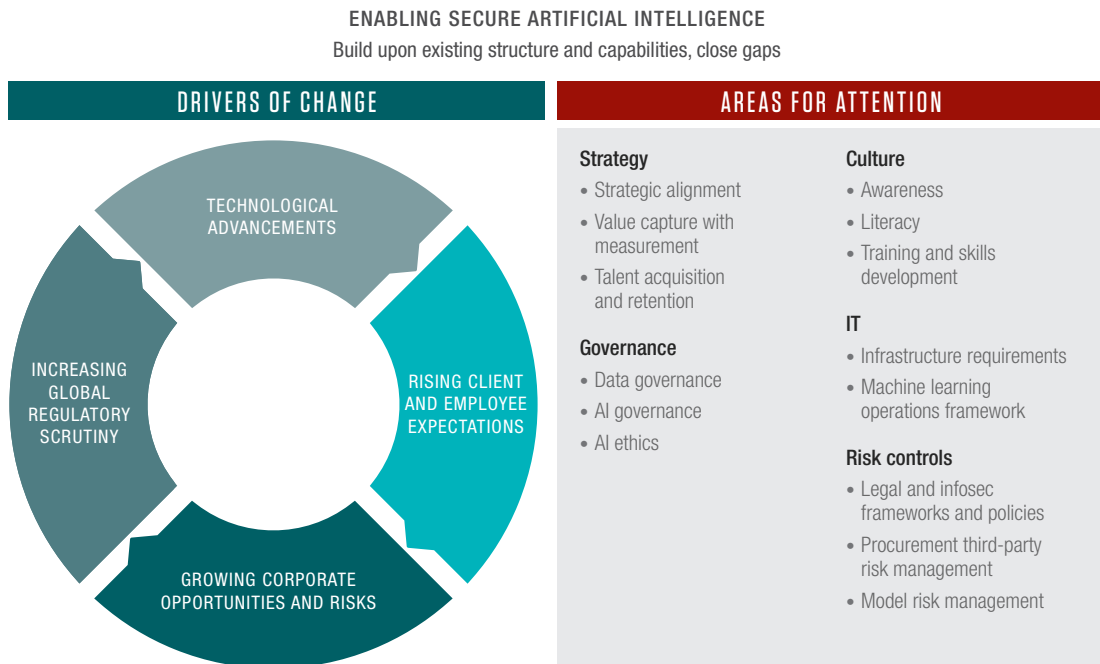
Given the growing importance of AI solutions, governing bodies around the globe, as well as technical experts and corporations, are trying to define guardrails and legislation to mitigate the risks associated with AI, balanced with innovation and the benefits the solutions bring. This is the exact same goal mentioned previously regarding data governance: it strives to achieve an effective way of balancing risk control with user-enabling innovation and insight generation. There is general consensus around the broad areas that require

Figure 6: Ethics guidelines



Source: Adapted from the E.U.'s Guidelines for Trustworthy AI

Figure 7: Keeping the end-to-end view in mind when building an AI governance framework



Source: Bank Julius Baer

attention, with a spotlight on ethical considerations around topics such as fairness, bias, and explainability – topics that are also covered under data ethics.

The E.U., to name just one example, proposed in their Guidelines for Trustworthy AI, that all AI solutions be lawful, ethical, and robust technologically and socially (Figure 6). Regarding ethics, it specifies four ethics principles: respect for human autonomy, prevention of harm, fairness, and explicability. It also suggests seven requirements to realize these principles: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental wellbeing; and accountability [AI HLEG (2019)].

For organizations, establishing AI governance to meet the requirements of regulators and legislators globally is critically important for ensuring that implemented solutions will withstand the test of time, and will not fail to evolve alongside regulatory requirements. A recent study by Ernst & Young found that while organizations and regulatory bodies broadly agree on the areas of focus for trustworthy AI, the importance of the individual principles is weighted differently [EY (2023)]. In addition, the regulatory landscape is still in flux, so the full scope of final legislative requirements cannot fully be judged yet.

As learned on the data governance journey, to govern AI efficiently within organizations requires a cross functional approach [Schneider et al. (2023)]. In addition to basic governance steps, such as defining principles for good model development and specifying AI principles that align with corporate values, experts from different domains need to collaborate make AI governance useful across a model's lifecycle (Figure 7).

Complementing the data governance experts, who curate and can help identify high-quality data sources, experts from the legal, information security, and IT domains are needed to ensure the proper operation of models. Data scientists and model risk managers need to monitor and validate model performance throughout the model lifecycle. Beyond the operational, governance bodies need to be established or upskilled to check for ethical considerations and risks associated with models [Blackman (2022)]. In addition, policies and controls need to be updated, or newly created, to address AI-related risks and to provide guidance to those working with the models. It is essential that this does not create additional overhead for data users, who face pressure from business management to provide information or solutions fast.

Applying what we learnt from data governance 1.0, we cannot start by only looking at the models and risks, we must include the consumer and business perspectives, and leverage technology as an enabler. Based on industry experience, only 10-20% of AI ideas and early proof of concepts actually make it all the way to production. Taking the right AI governance steps early in the process can increase the chances for success.

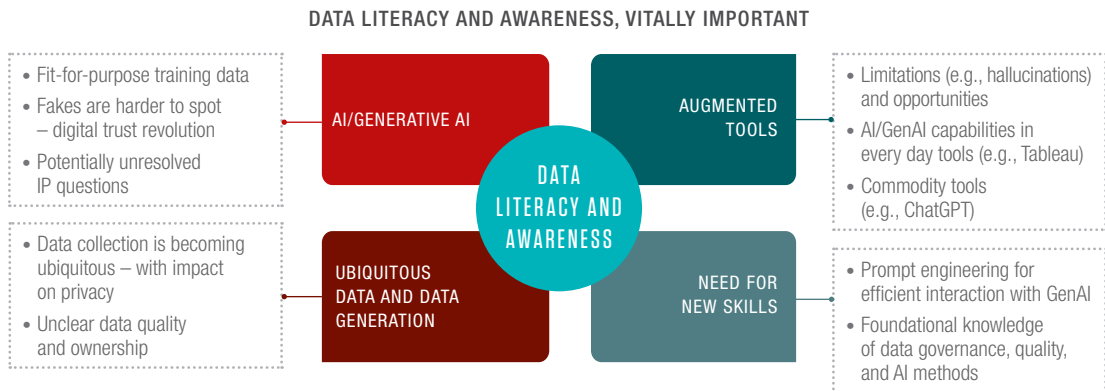
A number of points that could be considered at the ideation stage of a potential use-case include:

- **Strategy:** discuss if the use-case is really a direction you want to go as a firm. What would be the unintended outcomes if successful (impact on people, clients, and risk) and decide whether to park or move forward.
- **Governance:** could the use-case cause ethical issues or lead to a negative impact on society. Do you have the data to support the use-case, or only "some of it", which means outcomes will not be representative.
- **Culture:** do the people working on the use-case have the right skills and know the questions they should be asking from both ethical and risk perspectives.
- **IT:** can the data the team will be using legally be used in the environment they are looking to experiment in. Are they building in an environment that could never be productionalized.
- **Risk and controls:** does the team have the rights to use the input or output data, or could intellectual property be violated, or could data from third-parties be used in a way that would violate contracts.

A lightweight AI governance tollgate process at the idea stage can help avoid wasting time and resources on use-cases that will never go to production. Experimentation and exploration are nevertheless important, but should be understood by the business to be exactly that to set expectations.

Similar to the concept of security by design in software engineering, and the adoption of technology to enable data governance 3.0, AI can help AI governance, especially when it comes to managing the risks associated with the models themselves.

AI does not only support the identification of models across the organization. Embedding machine readable controls at the different stages of the model development lifecycle into the models themselves, unlocks the capability to validate models against different regulations, as well as continuously testing the models once in production to monitor their performance

Figure 8: Areas where data literacy and awareness are particularly critical

Source: Bank Julius Baer

and to ensure they do not “drift” into a state that is no longer within the expected risk tolerances [Aristi Baquero et al. (2020)]. The controls themselves become part of the model, embedded in place during model versioning, testing, and release cycles, and can be enhanced with further controls, as regulation and societal needs change.

Products are emerging that provide sets of machine-readable controls that represent different regulations or standards, and these controls can be downloaded and embedded into the models themselves as part of the model development lifecycle. Many of the vendors who are providing pre-built models/AI capabilities are being pressured to provide evidence of compliance with regulatory demands. With the regulatory space evolving so rapidly, creating the controls in a way that they can easily be embedded into models will benefit organizations that are “training their own” models, and vendors who can build in the controls as part of the offered solutions. It will be extremely difficult for enterprises to scale their use of AI without embedding controls as part of the model design.

The launch of ChatGPT in November 2022 brought GenAI and its capabilities to the forefront of public attention. Hailed as a tipping point for AI just two weeks after its launch [Mollick (2022)], McKinsey & Co. estimate the value potential of GenAI to be between U.S.\$2.6 trillion and U.S.\$4.4 trillion annually [McKinsey (2023b)]. GenAI solutions are usually built on foundation models, which are “pre-trained on large, unlabeled datasets and capable of a wide array of applications [...] and can then be fine-tuned for specific tasks” [IBM (n.d.)].

From a governance perspective, foundational models introduce another layer of complexity that can be addressed across three broad categories:

- Foundation models and commercial applications are usually trained outside an organization’s control, so there is no understanding as to whether the data used for training is representative and legally allowed to be used [Bommasani et al. (2023), Heaven (2023)], for example with regards to the use of plagiarized content or content that is created on copyrighted materials. Companies offering foundation models only publish select information and justify the lack of transparency with the protection of trade secrets, as well as the risk of bad actors gaming or hijacking the models [OpenAI (2023)].
- LLMs are statistical models at their core and come with certain limitations that are impactful, but the limits are not well understood. This is especially apparent in GenAI with hallucinated answers, which are statistically probable, but completely untrue.
- With the promise and popularity of GenAI, many vendors are adding components to their offerings (e.g., copilots), which increases the pressure on third-party risk management and technology providers to stay on top of these “features” being released into existing tools (some providers, such as Microsoft, offer to shield users of their models from possible lawsuits).

Regulators and organizations alike are keen to capitalize on the benefits that GenAI solutions offer but are also trying to understand and guardrail its specific risks. While the E.U. has included foundation models in its risk-based approach to AI regulation [European Council (2023)], the risk-profile is still evolving, which will add further turbulence to the AI regulatory space.

Outside of the risks mentioned above, the most important topic to cover when considering a corporate governance approach to GenAI, is awareness building with end-users coupled with data literacy (Figure 8). They need to understand the limitations (e.g., hallucinations, plagiarism, etc.) and the risks (potential loss of personal data) of foundation-model-based solutions, alongside their accountability (e.g., checking the correctness of content). Upskilling on how to most effectively interact with the solutions (e.g., prompt engineering) can help drive user-developed solutions for the areas they are the experts in, while understanding the risks involved.

5. CONCLUSION: APPLYING DATA GOVERNANCE LESSONS TO AI GOVERNANCE

The evolution of data governance taught us key lessons that we can now leverage as enterprise AI governance matures. From concepts through to actionable metadata linked to physical data, we learned that technological advancements far outpace our ability to govern through traditional methods.

Similar to “security by design”, which we see embedded in software engineering around the globe, governance needs to be built-in as part of the design and become a natural part of the ecosystem. In the case of data, the physical data assets themselves need to contain the metadata that enables the identification of risk, privacy, security, quality, and usability aspects of that data to enhance business and shareholder value. In the case of AI governance, the AI governance controls need be incorporated as part of the code of the model, generating the artifacts and evidence needed for model validation, trustworthiness, and ongoing monitoring.

The capabilities of LLMs will enable faster evolution of regulation and expectations on reporting. Today, regulatory and governing bodies work in the world of analogue rules and principles that are open to interpretation when being implemented by the organizations in scope. Since 2018, we have seen a number of regulatory bodies explore a more digital machine-readable approach to rules and regulation [PwC (2021), Ledger and McGill (2023)], which I expect will be further enabled through the strengths that LLMs have to turn unstructured non-digital content into machine-readable content.

All the data modeling work that regulated companies have undertaken to meet data regulations would reap even greater benefits, given that digital regulatory reporting (DRR) requires common data models to be effective and to ensure all parties

are “speaking” the same language. This is just one example where the evolution from data governance 1.0 to 3.0 shows us that we need to create the building blocks of the future today.

Organizations that have not linked data governance to physical data, or have not captured the needed metadata, or not modeled the data, are going to have a much harder time trying to meet future demands while generating business and shareholder value. Real-time financial and regulatory reporting may have sounded like an unachievable goal ten years ago, and it may be another ten-plus years before it becomes a reality, but it is certainly something that companies need to be creating the foundation for.

With the fast evolution of AI regulation taking shape around the world [IAPP (2023)], it is obvious that the “global reckoning on AI governance” is coming in the not-too-distant future, where we are seeing an initial divergence on a global scale (similar to the divergence that took place on data protection when the GDPR was introduced in Europe). Certain countries/geo-political alliances will take on more risk, regulate less, trying to leverage the capabilities of AI to upskill populations and improve economic conditions. On the other extreme, we have the heavily regulated E.U., which will struggle to innovate under the burden of expansive regulation [Greenacre (2023), Jorge Ricart and Alvarez-Aragones (2023)].

Converting regulation into machine-readable control frameworks, which can be modified, enhanced, and added to, enables the controls to mature and shape alongside the regulation. It is key to embed these controls as part of the AI development lifecycle, so as the controls change, they can easily be applied to both existing and new AI models. For example, you may have an E.U. AI Act set of controls as code, which can be called from different points in the AI development lifecycle and post-production for ongoing monitoring. This is not a new concept. In data lifecycle management there are several machine-readable controls that are applied at different stages, from data privacy classification to when data can be erased – pieces of callable code, ranging from standard scripts to running a machine learning algorithm at a certain time in the data lifecycle – a data governance 3.0 lesson we can leverage to take AI governance from infancy to a value-generating, scalable asset.

I will close with another Peter Drucker quote: “The relevant question is not simply what shall we do tomorrow, but rather what shall we do today in order to get ready for tomorrow” [Power (2018)].

REFERENCES

- Abdullahi, A., 2023, "Five ways to improve the governance of unstructured data," TechRepublic, August 28, <http://tinyurl.com/yc5vy46j>
- AI HLEG, 2019, "Ethics guidelines for trustworthy AI," High-Level Expert Group on Artificial Intelligence, April 8, <http://tinyurl.com/nyenub54>
- Appel, G., J. Neelbauer, and D. A. Schweidel, 2023, "Generative AI has an intellectual property problem," Harvard Business Review, April 7, <http://tinyurl.com/234z7jn2>
- Aristi Baquero, J., R. Burkhardt, A. Govindarajan, and T. Wallace, 2020, "Derisking AI by design: how to build risk management into AI development," McKinsey & Co., August 13, <http://tinyurl.com/5488ahyc>
- Awati, R., 2023, "Garbage in, garbage out (GIGO)," TechTarget, June, <http://tinyurl.com/ms6pyay7>
- Birkstedt, T., M. Minkinen, T. Anushree, and M. Mäntymäki, 2023, "AI governance: themes, knowledge gaps and future agendas," Internet Research 33:7, 133-167
- Blackman, R., 2022, "Why you need an AI ethics committee," Harvard Business Review, July-August, <http://tinyurl.com/yanv59cu>
- Bommasani, R., K. Klyman, S. Longpre, S. Kapoor, N. Maslej, B. Xiong, D. Zhang, and P. Liang, 2023, "The Foundation Model Transparency Index," Stanford University
- Brous, P., M. Janssen, and R. Krans, 2020, Data governance as success factor for data science. Responsible design, implementation and use of information and communication technology, Springer
- DAMA International, 2017, DAMA – DMBOK. Data management body of knowledge, Technics Publications
- Deighton, A., 2023, "Three reasons why your organization needs a data product strategy," Forbes, February 10, <http://tinyurl.com/yx6f9awu>
- Desai, V., T. Fountaine, and K. Rowshankish, 2022, "A better way to put your data to work," Harvard Business Review, July-August, <http://tinyurl.com/33b7h8ap>
- Drenik, G., 2023, "Data quality for good AI outcomes," Forbes, August 15, <http://tinyurl.com/v77nrxm>
- E.C., 2020, "Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment," European Commission, July 17, <http://tinyurl.com/3szc5aen>
- European Council, 2023, "Artificial Intelligence Act: Council and Parliament strike a deal on the first rules for AI in the world," December 9, <http://tinyurl.com/2s36kpkp>
- European Parliament, 2023, "EU AI Act: first regulation on artificial intelligence," December 19, <http://tinyurl.com/bdkkb4r>
- E.U., 2018, The General Data Protection Regulation applies in all Member States from 25 May 2018," European Union, May 24, <http://tinyurl.com/muybvc3p>
- EY, 2023, "The artificial intelligence (AI) global regulatory landscape," Ernst & Young, <http://tinyurl.com/y377fxf>
- Famularo, A., 2019, "The evolution of data governance," Forbes, March 11, <http://tinyurl.com/5xxcbxuy>
- Fowler, M., 2019, "How to move beyond a monolithic data lake to a distributed data mesh," martinofowler.com, May 20, <http://tinyurl.com/sft7e8sh>
- Greenacre, M., 2023, "EU/US divergence in data protection holds lessons for global regulation of artificial intelligence, experts say," Science Business, September 28, <http://tinyurl.com/mrxtyep>
- Hasselbalch, G., and P. Tranberg, 2016, Data ethics – the new competitive advantage, Publishare
- Heaven, W. D., 2023, "The inside story of how ChatGPT was built from the people who made it," Technology Review, March 3, <http://tinyurl.com/2vh7ma9r>
- Hinkle, O., 2020, "The evolution of data governance," Dataversity, May 18, <http://tinyurl.com/5eahdd8c>
- HKMA, 2019, "Reshaping banking with artificial intelligence," Hong Kong Monetary Authority and PwC, <http://tinyurl.com/4fpn2kxb>
- IAPP, 2023, "Global AI legislation tracker," International Association of Privacy Professionals, <http://tinyurl.com/mwum42j7>
- IBM, n.d. "What is an AI model?" <http://tinyurl.com/4us7xcf8>
- Jorge Ricart, R., and P. Alvarez-Aragones, 2023, "The geopolitics of generative AI: international implications and the role of the European Union," Elicano Royal Institute, November 27, <http://tinyurl.com/2s42wb8c>
- Ledger, M.-A., and A. McGill, 2023, "Digital regulatory reporting. The evolution of global initiatives," Bank of England, <http://tinyurl.com/mr238drk>
- Loftus, T., 2018, "The morning download: Facebook at center of global reckoning on data governance," Wall Street Journal, March 19, <http://tinyurl.com/2mz2szwy>
- MAS, 2023, "Veritas initiative," Monetary Authority of Singapore, October 26, <http://tinyurl.com/2t5b2hsp>
- McConnell, J., 2020, "Leadership everywhere means reversed leadership," Global Peter Drucker Forum, September 30, <http://tinyurl.com/yckv9mw>
- McKinsey, 2020, "Designing data governance that delivers value," McKinsey & Co., June 26, <http://tinyurl.com/4rmj9fw2>
- McKinsey, 2022, "How to unlock the full value of data? Manage it like a product," McKinsey & Co., June 14, <http://tinyurl.com/5n6wxjtx>
- McKinsey, 2023a, "Demystifying data mesh," McKinsey & Co., June 28, <http://tinyurl.com/bdfxrmja>
- McKinsey, 2023b, "The economic potential of generative AI: the next productivity frontier," McKinsey & Co., June 14, <http://tinyurl.com/3ypwe2y5>
- Mollick, E., 2022, "ChatGPT is a tipping point for AI," Harvard Business Review, December 14, <http://tinyurl.com/mskfh6pt>
- ODI, 2021, "Data ethics canvas," June 28, Open Data Institute, <http://tinyurl.com/2s3ww7a7>
- OECD, 2023, "G7 Hiroshima Process on generative artificial intelligence (AI)," September 7, <http://tinyurl.com/yck6jicz>
- OpenAI., 2023, "GPT-4 technical report," <http://tinyurl.com/4e7ez627>
- Power, D. J., 2018, "What are Drucker's views on planning and decision making?" DSSResources, April 23, <http://tinyurl.com/2hvay9yw>
- PwC, 2021, "Risk and regulatory outlook 2021 – key developments in Southeast Asia – digitalising regulatory reporting," <http://tinyurl.com/k27ccpsk>
- Rosencrance, L., 2024, "Five strategies for managing unstructured data," Techopedia, January 16, <http://tinyurl.com/3kwx9hmv>
- Schneider, J., R. Abraham, C. Meske, and J. vom Brocke, 2023, "Artificial intelligence governance for businesses," Information Systems Management 40:3, 229-249
- Talagala, N., 2022, "Data as the new oil is not enough: four principles for avoiding data fires," Forbes, May 2, <http://tinyurl.com/4zjzza7d>
- Violino, B., 2023, "Eight tips for unleashing the power of unstructured data," CIO Magazine, November 28, <http://tinyurl.com/bde7hb42>
- Watts, M., 2021, "Why data is the new oil," Futurescot, November 17, <http://tinyurl.com/39k5ifs4>
- Woods, D., 2016, "Why data lakes are evil," Forbes, August 26, <http://tinyurl.com/4az9h2hp>

“DATA ENTREPRENEURS OF THE WORLD, UNITE!” HOW BUSINESS LEADERS SHOULD REACT TO THE EMERGENCE OF DATA COOPERATIVES

JOSÉ PARRA-MOYANO | Professor of Digital Strategy, IMD

ABSTRACT

Data cooperatives, entities that allow individuals to pool together their personal data to gain collective bargaining power and enable them to monetize their data, are emerging. This article describes the economic mechanisms that motivate the emergence of data cooperatives and analyzes the challenges and opportunities that the existence of these cooperatives implies for business leaders.

1. INTRODUCTION

Over the past few decades, the emergence of social media and digital platforms has catalyzed an exponential increase in the creation and accumulation of personal data. This surge originates predominantly from the seamless integration of these platforms into daily life as they capture a myriad of data points ranging from consumer behaviors and preferences to social interactions and personal interests. The extensive and diverse nature of this data has made it a treasure trove for businesses aiming to derive value, innovate, and gain a competitive edge.

Because of that, data has become a factor of production, just like capital and labor (i.e., an input that is required for companies to develop and market their products and services). This fact has been further magnified by the “artificial intelligence” (AI) and data science revolution. AI and “machine learning” (ML) technologies thrive on large datasets, often revealing valuable insights that can optimize operations, predict market trends, personalize customer experiences, and drive strategic decisions, thereby cementing data’s role as a foundational element of modern economic production.

2. DATA IS ECONOMICALLY UNDERUTILIZED

Interestingly, personal data is being underutilized. To understand why, we need to recognize that data as a factor of production differs significantly from traditional factors of production like labor and capital. It is unique because of its multi-user nature: data can be used by multiple entities at the same time without being depleted. For example, a single dataset about consumer preferences can simultaneously benefit an advertising firm, a market research company, and a product development team, thus implying that a person’s personal data can create value for several organizations at the same time.

Setting aside two issues for a short moment, namely, who should earn the profits from the analysis of personal data and what are the privacy implications of an increased analysis of personal data by many organizations simultaneously, from a purely economic perspective, having personal data utilized by several companies at the same time would be economically beneficial. The multi-user nature of this (new) factor of production would create a parallelization of the value-generation process resulting in increased value for organizations.

3. INCREASED VALUE REQUIRES VOLUME

Since data belongs to the people to whom it refers, it would be quite natural for individuals to join associations, cooperatives, or companies that would gather their data from various sources and then offer this data to other organizations in exchange for a fee. Individuals would need to pool their data together to create a large enough dataset so that the data analysis will result in the valuable insights to which we refer.

4. DATA COOPERATIVES: POTENTIAL TO RESHUFFLE THE DECK IN THE DIGITAL ECONOMY

Amidst this backdrop, data cooperatives have emerged as a revolutionary concept. These entities allow individuals to pool their personal data, thus creating collective bargaining power. This pooling enables individuals to monetize their data, allowing entities other than the platforms on which the data was created to access and derive value from it.

Examples of data cooperatives include Swash (swashapp.io), MIDATA (midata.coop), Driver's Seat (driversseat.coop), SalusCoop (saluscoop.org) and the Data Worker's Union (dataworkers.org). Swash was conceived as a way to enable users to earn income from their browsing data and it offers a simple yet effective way for individuals surfing the internet to gain from the digital footprint they leave. Meanwhile, MIDATA focuses on health data, creating opportunities for individuals to contribute to medical research and healthcare improvements while maintaining control over their personal health information. Also in the area of healthcare, SalusCoop operates on the principle of voluntary, non-profit sharing of health information by its members. The cooperative emphasizes the ethical use of this data for research purposes, ensuring that the data is used to benefit individual and public health outcomes while respecting the privacy and rights of the data providers.

Driver's Seat caters to gig economy drivers, providing them with insights and tools to better manage and benefit from their work-related data. Finally, the Data Worker's Union advocates for the rights of data producers across various sectors, emphasizing fair treatment and use of data.

Each cooperative offers unique ways for individuals to monetize and control their data, reshaping how personal data is viewed and utilized in the digital economy.

4.1 Learning from the past to better understand the future

In the late 18th century, Adam Smith helped define the concept of labor as a factor of production. This new concept then contributed to the emergence of the labor unions shortly after, which were formalized and legally recognized by the early 19th century.

Interestingly, the notion that data is a factor of production only emerged over the last 10 to 15 years. There is an analogy to be made between data cooperatives and labor unions in the context of data as a form of digital labor. Just as labor unions collectivized the workforce to negotiate better terms and protections for workers, data cooperatives aggregate individual data contributions in the hope that citizens could participate in the profits that emerge from the analysis of their data. This aggregation strengthens the negotiating power of individuals over their data, much like unions do for labor rights and wages. In unions, the collective bargaining power helps to secure fairer terms of employment; similarly, in data cooperatives, this collective strength ensures better control and potential monetization of personal data. Both institutions serve to balance power dynamics – unions between workers and employers, and data cooperatives between individual data providers and data-using entities.

The key difference between data cooperatives and labor unions is that data cooperatives facilitate the creation of new wealth (rather than a distribution) and empower citizens to become "data entrepreneurs" that monetize a "product" they own. In this sense, data cooperatives help liberalize data as a factor of production, increase competition, and enable citizens to participate in the free market.

4.2 Possible consequences of the adoption of data cooperatives

The emergence of data cooperatives may mark a pivotal shift in the digital economy, resulting in a more efficient utilization of data, the democratization of its monetary benefits, and the creation of novel income streams in an AI-transformed society.

- **Efficient utilization of data in the economy:** data cooperatives streamline the aggregation and application of data. By pooling data from numerous individuals, these cooperatives amass a rich, diverse dataset that is more reflective of the broader population. This aggregation

“

Data cooperatives can offer a novel approach to managing privacy concerns in the digital age by embracing the concept of “sending the algorithm to the data”.

”

enhances the data's utility for a range of applications, from healthcare research to consumer behavior analysis. Such comprehensive datasets are invaluable for training AI models, leading to more accurate and effective AI solutions. As a result, the economy benefits from a more precise understanding of trends, behaviors, and needs, driving innovation and progress across various sectors.

- **Monetization and passive income for citizens:** data cooperatives empower individuals to monetize their personal data. In an era in which data is a critical asset, individuals contributing to these cooperatives can receive compensation for their data, creating a new stream of passive income. This model offers a unique opportunity for individuals to capitalize on the digital footprints they naturally create, turning a routine activity into a financial asset. This not only provides economic benefits to the data contributors but also encourages a more equitable distribution of the wealth generated from data-driven activities.
- **Supporting retirement and alternative income sources:** the passive income generated through data cooperatives can be a significant support for individuals, especially during retirement. As traditional retirement funds face challenges, the additional income from data monetization can provide a much-needed financial buffer. Also, this approach is particularly potent in offsetting the economic impacts of AI and automation, which may displace traditional jobs. By monetizing their data, individuals can supplement their income, providing financial resilience in a rapidly changing job market.

Consequently, data cooperatives may represent a transformative approach to data management and utilization in the digital age. They facilitate efficient data use across industries, enable individuals to benefit financially from their personal data, and offer a novel solution to some of the economic challenges posed by AI and technological advancements. This model heralds a new era of data democracy, where the value generated from data is shared more broadly across society.

4.3 Privacy-preserving data cooperatives

Naturally, one concern that emerges with the rise of data cooperatives is privacy. It is easy to see how citizens' privacy would be (even more) depleted when an additional number of firms are able to access their data. However, there do exist privacy-preserving techniques that enable the conditional use of data so that insights from data can emerge in a privacy-preserving manner. Data cooperatives can offer a novel approach to managing privacy concerns in the digital age by embracing the concept of “sending the algorithm to the data”. This method ensures that personal data remains within the cooperative's secure environment while still being useful for external AI applications.

The core of this approach lies in the cooperative receiving queries or AI models from third-parties. Instead of transferring the data out, the cooperative runs the analysis or trains the AI within its secure system. This allows the AI to learn from the data or for queries to be answered without the data ever leaving the cooperative's secure environment. And what is more important, the members of the cooperative can approve or decline the use of their data on a case-by-case basis. Hence, if very few members of the cooperative agree to use their data for a particular purpose at a given price, then the price to access that data would increase. This would result in a de-facto market-driven pricing structure that would reveal the actual value of data.

Additionally, this approach has several advantages:

- **Enhanced data security:** since the data never leaves the cooperative, the risk of breaches and unauthorized access is significantly reduced. This is crucial, especially for sensitive data like health or financial information.

- **Compliance with privacy regulations:** this method aligns with global data privacy regulations like GDPR, which emphasize data minimization and the principle of processing data within the entity that owns it.
- **Maintaining data integrity:** by keeping data within the cooperative, the integrity of the data is maintained. There is less risk of data being tampered with or mishandled when it is not being transferred across different platforms or networks.

In practical terms, a health data cooperative could receive an AI model designed to identify patterns in medical imaging. Instead of sending out the medical images, the cooperative would run the AI model on its internal servers. The AI learns from the data, but the data itself remains securely within the cooperative. Only the insights or results from the AI analysis are then shared with the external party.

4.4 Challenges

Data cooperatives, employing the concept of “sending the algorithm to the data”, offer a forward-thinking resolution to privacy concerns in AI and data analytics. However, this model also involves its own set of challenges that would need to be addressed before its effective implementation.

- **Technical complexity:** the infrastructure required to handle sophisticated AI models and queries in-house is substantial. Cooperatives must invest in high-powered computing resources and develop robust data processing frameworks. Ensuring the seamless integration of external AI models with internal systems can also be technically challenging.
- **Ensuring AI model security:** AI models sent to cooperatives for training could potentially be designed in ways that extract or infer sensitive information. Rigorous evaluation and testing of these models for privacy compliance are critical.
- **Data quality assurance:** as cooperatives aggregate data from various sources, ensuring the consistency, accuracy, and quality of this data becomes essential. Poor data quality can lead to inaccurate AI training and unreliable results.
- **Scalability issues:** as the amount of data and the complexity of AI models increase, scaling the infrastructure while maintaining data privacy and processing efficiency can be challenging for cooperatives.



- **Legal and regulatory compliance:** navigating the evolving landscape of data privacy laws and ensuring compliance with multiple jurisdictions' regulations is a complex task, requiring constant vigilance and adaptation.

In addressing these challenges, data cooperatives need to develop comprehensive strategies that include investing in technology, training personnel, and establishing robust data governance frameworks. This involves not only technological investments but also fostering a culture of data privacy and security within the cooperative.

5. IMPLICATIONS FOR BUSINESS LEADERS

The increasing emergence of data cooperatives has significant implications for leaders in the financial services sector, offering many opportunities as well as specific challenges. Understanding and embracing these changes is crucial for staying competitive and innovative.

- **Insights into consumer behavior:** data cooperatives provide access to rich, diversified consumer data. Financial leaders can gain deeper insights into customer behaviors, preferences, and needs. This data can inform product development, marketing strategies, and customer service improvements, leading to more tailored financial services.
- **Enhanced risk assessment:** the detailed data from cooperatives can improve risk assessment models. By accessing more comprehensive datasets, financial institutions can refine their credit scoring systems, detect fraud more effectively, and manage risks better.
- **Regulatory compliance:** data cooperatives operate within stringent privacy and data protection frameworks. Financial leaders can leverage these cooperatives to ensure compliance with regulations like GDPR while utilizing essential data for business operations.
- **New business models:** the cooperative model opens avenues for new business models. Financial institutions can collaborate with these cooperatives, offering financial services tailored to the cooperative members, such as loans, insurance, or investment products based on the aggregated data.

- **Competitive advantage:** early adopters of this model in the financial services sector could gain a significant competitive edge. By accessing a broader range of data, financial institutions can offer more personalized services, enhancing customer satisfaction and loyalty.
- **Data-driven innovation:** the cooperative model encourages innovation. Leaders of financial services organizations can use the diverse data to develop new financial products and services, leveraging AI and machine learning for better financial forecasting and decision making.
- **Building customer trust:** by partnering with data cooperatives that prioritize data privacy and user control, financial institutions can build greater trust with their customers. This approach demonstrates a commitment to ethical data use and customer-centric practices.
- **Strategic partnerships:** the financial services sector can establish strategic partnerships with data cooperatives. These partnerships can lead to shared initiatives, joint ventures, or co-developed financial products, benefiting both parties.

Despite these many opportunities, adapting to this new model requires overcoming certain challenges as well, including integrating cooperative data with existing systems, ensuring data security, and navigating the cooperative's governance structure.

For leaders of financial services organizations, the rise of data cooperatives is not just a trend to observe but a strategic opportunity to harness. By understanding and integrating this model into their data strategy, financial institutions can enhance their services, innovate more effectively, and build stronger customer relationships in the data-driven era.

6. CONCLUSION

The potential of data cooperatives to transform how data is utilized in the economy is immense. By offering a fairer use of personal data, they pave the way for more innovative and personalized AI-driven solutions across industries. The model they propose harmonizes the need for data-driven insights with the critical importance of maintaining individual privacy.

Moreover, the successful implementation of data cooperatives could lead to a more equitable digital economy. By enabling individuals to monetize their data and become data entrepreneurs, these cooperatives provide a means for people to benefit directly from the digital economy, potentially offsetting job losses in other sectors due to AI and automation.¹

The data cooperative model, therefore, is not just a step towards a more inclusive and balanced digital future, but also a potential solution to the loss of work caused by the increasing implementation of AI. As this model gains traction, it could set a new standard for data handling and utilization, fostering a more competitive and diverse market and empowering individuals as key stakeholders in the data economy. This shift promises to catalyze innovation while upholding the principles of privacy and ethical data use, heralding a new era in the data-driven digital economy.

¹ Ito, A., 2023, "The AI heretic," Business Insider, September 23, <http://tinyurl.com/27svrvmn>

REVOLUTIONIZING DATA GOVERNANCE FOR AI LARGE LANGUAGE MODELS

XAVIER LABRECQUE ST-VINCENT | Associate Partner, Capco

VARENYA PRASAD | Principal Consultant, Capco

ABSTRACT

In an artificial intelligence (AI) enabled organization, traditional data governance practices face challenges due to the complexity of AI algorithms, utilization of unstructured data, dynamic data transformations, integration with external data sources, and the lack of interpretability in AI models. To overcome these challenges, financial institutions can deploy strategies to increase transparency, refine metadata for unstructured data, and foster collaboration. Furthermore, data ownership and stewardship roles demand evolution in the AI-driven landscape. Ownership now encompasses AI models, algorithms, and insights. To address the needs of stakeholders and ensure responsible AI usage, collaboration, technical expertise, and a focus on governance and compliance become crucial. By adapting their data governance frameworks to accommodate the unique challenges presented by AI, financial institutions can maximize the value of AI technologies while maintaining data quality and trustworthiness. This transformation in data governance is essential for financial institutions to capitalize on the benefits of AI and maintain a competitive edge in the industry.

1. INTRODUCTION

Traditional data governance practices and risk maturity models are rendered obsolete or inadequate in the era of artificial intelligence (AI) adoption, particularly with the emergence of advanced technologies such as large language models (LLMs) like ChatGPT. Financial institutions are aggressively pursuing AI implementation to gain competitive advantage, aiming to optimize both employee and customer experiences. However, the rapid evolution of AI introduces significant challenges related to data privacy and transparency, necessitating a fundamental reboot of data governance within organizations. In response, we have developed a comprehensive framework tailored to address these concerns, enabling financial institutions to not only navigate the complexities of AI integration but also to capitalize on opportunities presented by “generative pre-trained transformers” (GPT) AI in specific use cases.

2. THE EARLY PROMISE: MAXIMIZING VALUE WITH AI

Generative AI's appeal has captivated many and we are now faced with a new paradigm in which traditional data governance practices and risk maturity frameworks have become obsolete or inadequate. The GenAI use cases underscore the empowerment of employees by facilitating rapid access to information, a task that traditionally consumed hours. This holds particular significance for business functions striving for efficiency gains and superior customer experiences. Moreover, AI offers the tantalizing prospect of driving innovation through dynamic machine learning algorithms.

Embracing comprehensive transformations is essential for ensuring readiness and to avoid being left behind in this fiercely competitive industry. However, it is imperative for organizations to recognize that outdated governance practices

and risk frameworks no longer suffice in this era of AI. Vigilant oversight remains crucial, yet advancements in AI-driven automation can streamline processes, reducing reliance on manual reviews. Organizations must prioritize adapting governance principles and frameworks to seamlessly integrate AI technologies.

3. THE CHALLENGE: REDEFINING DATA

In this AI-enabled world, the very essence of data has undergone a profound transformation, where “language is data” emerges as a cornerstone principle. Language, encompassing human speech, text, emotions, and sentiments, emerges as a primary source of input data, marking a departure from the conventional use of structured data. This shift is particularly pronounced in AI applications, notably those leveraging natural language processing, which thrive on the amalgamation of structured and unstructured data. Table 1 underscores the pivotal role of unstructured data in capturing nuanced and diverse information often overlooked by traditional structured datasets.

As datasets evolve into a blend of structured and unstructured data, ensuring data quality becomes paramount to its suitability for AI consumption. The suitability of data for AI ingestion critically depends on its quality, as it directly shapes the resulting outputs. Language analysis and sentiment classification necessitate the nuanced interpretation of human language, demanding governance frameworks to adapt accordingly. These frameworks must vigilantly monitor for corrupt tainted data, extract pertinent insights, and flag potential issues, underscoring the necessity for agile and robust governance structures in this AI-driven landscape.

“*In this AI-enabled world, the very essence of data has undergone a profound transformation, where “language is data” emerges as a cornerstone principle.*”

To tackle these challenges, our methodology helps ensure that risks are well understood and mitigated through the clear definition of use cases and establishment of guardrails for regulatory and audit purposes. For financial institutions to reap the benefits of GenAI, they need to focus on the most fundamental challenge of governing their data in an AI-enabled world.

3.1 Traditional data governance is no longer sufficient

With the rapid advancement of AI technology, we are witnessing a surge in new models and the creation of fresh data for various purposes. Though their inner workings may seem complex, their potential to revolutionize large institutions is undeniable. By addressing issues of data lineage, biases, and unintended consequences head-on, we enable a future where AI empowers us all.

The effectiveness of governing AI hinges on the data it uses and how transparent its uses/algorithms are. There are numerous challenges that are now present for financial institutions to overcome (Figure 1).

Table 1: Unstructured data eases the capture of rich and diverse information that traditional structured data might miss

	STRUCTURED DATA	UNSTRUCTURED DATA
Definition	<ul style="list-style-type: none"> • Data with a high degree of organization; follows a predefined model or schema. • Explicitly defined in columns and rows. 	<ul style="list-style-type: none"> • Data lacking a predefined data model; lacks a consistent structure. • Often required to convert raw data into a usable format.
Data sources	<ul style="list-style-type: none"> • Tabular data • Databases • Spreadsheets such as Excel, Google Sheets • Online forms 	<ul style="list-style-type: none"> • Word documents, PDF, emails, blogs, news articles, research journals, etc. • Images • Audios • Videos • Social media posts and commentary

Figure 1: Challenges to risk management and governance in an AI-enabled environment

1. COMPLEXITY OF AI DATA	2. TRANSPARENCY AND EXPLAINABILITY	3. DYNAMIC NATURE OF AI	4. REGULATORY COMPLIANCE	5. DATA QUALITY AND TRUST	6. DYNAMIC DATA ECOSYSTEM
Traditional governance frameworks struggle with the complexity and variety of AI data sources, which include unstructured and semi-structured data.	AI algorithms lack transparency and explainability compared to traditional statistical methods, raising concerns about bias and ethical implications of AI decisions.	Rapid advancements in AI technologies outpace the capabilities of traditional governance frameworks, leading to gaps and leaving organizations vulnerable to risks.	Evolving regulatory landscapes, especially concerning data privacy and ethical AI use, require governance frameworks to adapt accordingly.	Ensuring data quality and trustworthiness in AI-driven processes demands more sophisticated governance mechanisms to tackle AI data characteristics such as bias and data drift.	Data is constantly evolving, with new sources, formats, and volumes emerging rapidly. Traditional governance frameworks may struggle to keep pace leading to gaps in data governance coverage and effectiveness.

As AI becomes more prevalent, governance practices must adapt accordingly. We need to ensure the same level of oversight for AI data and models as we do for traditional data. This is essential to maintain high quality and trustworthiness, especially when it comes to the impact on clients. To meet this demand, data governance teams must step up and play a crucial role in navigating this new era of AI. Let us examine a couple examples to illustrate.

3.1.1 EXAMPLE 1: DATA LINEAGE IN AN AI-ENABLED ORGANIZATION

Data lineage, a fundamental aspect of traditional data governance, refers to the ability to track the origin, movement, and transformations of data throughout its lifecycle. In a traditional data environment, where data flows are relatively straightforward and well-defined, establishing data lineage is often feasible through manual documentation and tracking mechanisms.

However, when AI is introduced into the enterprise, traditional data lineage practices face significant challenges due to several reasons:

- **Complexity of AI algorithms:** AI algorithms, particularly deep learning models, constantly evolve, making it difficult to trace how data inputs are processed and transformed to produce outputs. Unlike traditional systems where data transformations are explicitly defined, AI algorithms learn and adapt based on complex patterns within the data, rendering manual lineage tracking impractical.

- **Unstructured data:** AI thrives on unstructured data such as text, images, and audio, which often lack clear lineage metadata. Traditional data lineage tools and techniques are designed for structured data, making them ill-equipped now.
- **Dynamic data transformations:** AI models continuously evolve and adapt as they ingest new data and learn from feedback. This dynamic nature of AI introduces a challenge as any previously captured lineage and data transformations will require at least constant updates, if not near real-time. Traditional lineage tools may struggle to keep pace with the rapid changes in AI models and data.
- **Integration with external data sources:** AI applications often rely on external data sources, such as third-party datasets and APIs, which may not provide comprehensive lineage information. Integrating external data sources into the enterprise data ecosystem introduces third-party governance, which further complicates the task of establishing end-to-end lineage.
- **Interpretability and explainability:** AI models are often characterized by their lack of interpretability and explainability, making it challenging to understand the rationale behind their decisions. Without clear visibility into how AI models utilize and transform data, establishing meaningful data lineage becomes elusive.

Table 2: AI considerations for data governance have a key role in defining the future operating model

FOCUS AREA	DESCRIPTION	AI CONSIDERATIONS FOR FINANCIAL INSTITUTIONS
Data governance team	A dedicated team responsible for overseeing the implementation and management of data governance initiatives. This team typically includes data stewards, data architects, and compliance officers.	Does your team have the right skillset to support the governance of GenAI models?
Data governance policy	A set of guidelines and rules that outline the principles, objectives, and responsibilities of data governance within the organization. This policy provides a framework for the implementation and enforcement of data governance practices.	How does the data governance policy address the specific considerations and challenges associated with AI, such as transparency, explainability, and bias mitigation?
Data stewardship	Data stewards are assigned to specific data domains or business areas and are responsible for ensuring the quality, integrity, and security of the data within their domain. They act as the custodians of the data and enforce data governance policies and standards.	How will data stewards work with AI teams to ensure that AI models are trained on quality data and that the outputs are accurate and reliable? Are stewards able to determine if the data is fit-for-purpose?
Data classification	The process of categorizing data is based on its sensitivity, criticality, and regulatory requirements. This classification helps determine the appropriate level of protection and access controls for different types of data.	How will data classification consider AI-specific requirements, such as identifying sensitive data used for training AI models or identifying data subject to regulatory compliance?
Data quality management	A set of processes and practices aimed at ensuring the accuracy, completeness, and consistency of data. This includes data profiling, data cleansing, and data validation activities to improve data quality.	How will data quality management address the unique challenges posed by AI, such as evolving data models and the need for ongoing monitoring and validation of AI model inputs and outputs? How will considerations of ethics and bias impact data quality measurements?
Metadata management	The management of metadata, which provides information about the data, including its structure, definitions, relationships, and lineage. Metadata management helps to improve data understanding and facilitates effective data governance.	How will metadata management capture and document the specific characteristics of AI models, including the algorithms, training data, and validation processes used? How will unstructured data be catalogued?
Data security and privacy	Policies, procedures, and controls to protect data from unauthorized access, breaches, and ensure compliance with privacy regulations. This includes access controls, encryption, data masking, and monitoring of data usage.	How will data security and privacy measures address the unique risks associated with AI, such as protecting sensitive training data and ensuring privacy in AI-driven decision-making processes?
Data governance tools	Software tools and technologies used to support and automate data governance activities. This includes data cataloguing tools, metadata management tools, data quality tools, and data lineage tools.	How will data governance tools integrate with AI platforms and technologies to enable effective management and oversight of AI models and their associated data?
Training and education	Ongoing training and education programs to raise awareness about data governance, promote best practices, and ensure that employees understand their roles and responsibilities in data governance.	How will training and education programs address the specific knowledge and skills required to understand and effectively govern AI technologies, including AI ethics, bias mitigation, and explainable AI?
Compliance and audit	Regular monitoring and audits to assess compliance with data governance policies and regulatory requirements. This includes internal and external audits to identify any gaps or non-compliance and taking any corrective actions.	How will compliance and audit processes assess the adherence to AI-specific regulatory requirements and ethical standards? Will audits include reviewing AI model development and decision-making processes?

To address these challenges, organizations can implement strategies to increase algorithmic transparency, enriching metadata for unstructured data, fostering cross-disciplinary collaboration, and continuously improving lineage practices. These approaches aim to enhance the understanding and tracking of data origin, movement, and transformations in AI environments.

3.1.2 EXAMPLE 2: DATA OWNERS AND STEWARDS ARE NOW CONSTRAINED TO GOVERN AI DATA

Traditionally, data ownership and stewardship roles are often assigned to specific departments or individuals responsible for managing and maintaining data assets within their respective domains. However, with AI, the scope of data ownership expands beyond traditional boundaries. Ownership now extends to encompass not only the raw data but also the AI models, algorithms, and derived insights generated from that data.

With the integration of AI into enterprise workflows, the roles and responsibilities associated with data ownership and stewardship must evolve to address the unique challenges and opportunities presented by AI technologies. In this AI-driven landscape, where language itself becomes a form of data, there emerges a wider range of stakeholders. This expansion of stakeholders necessitates not only retraining

but also broader engagement across the organization. This demands a holistic approach that emphasizes cross-functional collaboration, technical expertise, and a heightened focus on governance and compliance. The data governance team, therefore, needs to be retrained and upskilled to develop a deeper understanding of AI principles, algorithms, and techniques. The team will need to work closely with legal and compliance, data scientists, machine learning engineers, researchers, etc., to establish guidelines for responsible AI usage, monitor adherence to these guidelines, and address any ethical or legal concerns that may arise.

3.2 The solution: To embrace AI effectively, revolutionize your data governance

Data governance must boldly evolve across all its core functions alongside the rise of GenAI. As these tools and technologies advance, the expertise of subject matter specialists becomes essential in assessing data quality, enhancing their quality, and confirming their viability for algorithmic modeling.

To begin, financial institutions should assess current governance practices to identify strengths and weaknesses to understand how to align with strategic objectives. In this evolving landscape, the outlined considerations presented in Table 2 will need to be addressed by data leaders as they are key differentiators to enable organizations to perform their role to the best of their ability.



4. CONCLUSION

Traditional data governance frameworks face significant challenges when integrating AI technologies due to the complexities introduced by unstructured data, opaque algorithms, and dynamic data flows and transformations. As AI is embraced, strategies can be implemented to evolve data governance practices to ensure transparency, accountability, and ethical use of data.

Data ownership and stewardship roles also need to evolve in the AI-driven landscape. Ownership expands to include AI models, algorithms, and insights. Collaboration, technical

expertise, and a focus on governance and compliance are important to address the needs of stakeholders and ensure responsible and ethical AI usage.

To effectively embrace AI, financial institutions must revolutionize their data governance and start leveraging advanced tools and techniques for managing AI data. They should assess current practices, align them with objectives, and address key considerations like data quality, lineage, ownership, and compliance. This will help them navigate AI complexities and seize opportunities in specific use cases.

MUNICIPAL DATA ENGINES: COMMUNITY PRIVACY AND HOMELAND SECURITY

NICK REESE | Cofounder and COO, Frontier Foundry Corporation¹

ABSTRACT

Convergence is when two or more separate technologies are paired together to create a capability that is greater than the original technologies individually. The additional value of the converged system itself now opens new applications as well potentially new challenges. As policy conversations around emerging technology implications grow, the importance of considering convergence is paramount for effective and trustworthy implementation of technologies in municipal spaces. A connected community is not a technology but a convergence concept that touches millions of citizens, their privacy, and the critical infrastructure on which each of them depend. As with all examples of convergence, there are implications beyond the sum of their parts and connected communities is no exception. Officials and individual users are familiar with the implications of connected technologies on individual privacy but the concept of municipal, community, or regional privacy is new. The aggregated data of an entire community or region take the concept of privacy to the homeland security level, driving increased need for effective policies and controls to ensure the safety and security of citizens living inside these architectures. This article explores specific challenges for the implementation of municipal IoT and introduces the concept of privacy at the municipal, community, and regional levels.

1. INTRODUCTION

Emerging quickly and seemingly without warning, generative artificial intelligence (GenAI) reignited series of debates around governance, ethics, and technology proliferations and its impact on any number of aspects of the human condition from romantic relationships to human job loss to national security. For governments and policymakers, the topic of AI had been an area of general interest and discussion, but the introduction of ChatGPT in November of 2022 has accelerated debate and action. In the U.S, a new AI Executive Order was released by the Biden Administration [White House (2023)] and the European Union (E.U.) passed its AI Act [European Parliament (2023)]. While much of the debate around AI has thus far focused on specific models, ownership, output quality, security, or ethics, the issue of technology convergence has been largely absent from the discussion.

Convergence is when two or more separate technologies are paired together to create a capability that is greater than the original technologies individually. The additional value of the converged system itself now opens up new applications as well potentially new challenges. For example, unmanned aerial vehicles (UAV) or drones combine technologies that include computer optics, robotics, AI, telecommunications, aerospace technologies, and more. Alone, each of these technologies is significant but when paired together and aimed at a specific use, they form something completely new that is greater than any of the individual technologies that make it up. In the same way, convergence of other technologies is creating bigger challenges than the mere existence of GenAI tools. Convergence between cutting edge technologies like AI and quantum or outer space capabilities and AI have the potential to create far bigger impacts and should be addressed.

¹ The author holds a faculty position at New York University, where he teaches courses on Emerging Technology and National Security and on Connected Communities. He is a member of the Homeland Security Advisory Board at George Washington University and is the former Director of Emerging Technology Policy at the U.S. Department of Homeland Security.

A trap when talking about technology concepts is to keep them overly abstract. Talking about AI as a general concept leads to abstraction that borders on uselessness. The same can be true when talking about convergence. It is a generally easy concept but without a real use case, it can feel like much less of a factor than it is. Rather than discuss convergence as a concept, this article will use the application of convergence in municipal environments as a way to properly convey the message and the challenges.

Known as “smart cities” or “connected communities”, connected technology deployments in municipal environments, rural and urban, is growing. Citizen demand for such technologies is also growing as potential solutions for traffic problems, energy use, and resource distribution, among others, are proposed. There are few technology architectures that impact more people more directly than a connected community deployment in a municipal environment of any size. Internet connected devices that monitor and optimize our resource distribution also create cyber vulnerabilities where none previously existed. The study of critical infrastructure risk and dependence has been ongoing for years but the addition of potentially tens of thousands of connected devices to critical infrastructure without a standard method of deployment or security requirements renders most of the cyber risk assessments void. Technology convergence is becoming a serious potential threat to our homeland security and our ability to provide critical services, and it impacts more people directly, and through their data privacy concerns, than any technology individually.

In this article, we will explore what a connected community is, what technology comprises its architecture, and discuss the gaps we see as these architectures continue to be developed and deployed on top of critical infrastructure. We will explore privacy issues, not at the individual level but at the municipal level, and show how municipal privacy extends to a homeland security issue rather than a law enforcement issue. Finally, we will discuss the need for new risk models, powered by AI, and for interoperability of connected community technologies. Technology convergence is an issue that will touch everyone, but no single use case will touch as many as connected communities.

2. WHAT IS A CONNECTED COMMUNITY

In a 2020 literature review, multiple authors define the term “smart cities” as generally referring to the use of technology-based solutions to enhance the quality of life for citizens, improve interactions with government, and promote sustainable development [Ismagilova et al. (2020)]. A smart city, or connected community, is not itself a technology, rather it is a concept and a perfect example of convergence. A connected community seeks to bring deployed technologies to bear against problems in municipal environments. The specific problems that are targeted for solution depend heavily on the municipality itself. For example, a rural community may choose to incorporate a smart irrigation system into its architecture while urban environments may choose to focus on traffic issues or WiFi in public spaces. On some levels, a connected community architecture must function this way because the implementation of technology in a municipal environment must directly reflect the needs and realities of the municipality in question. What works for Pittsburgh may not work for Seattle because of the different needs and environments of each city. In all cases, architectures bring some combination of the following technologies to form a foundation that seeks to solve a given set of municipal problems:

- internet of things (IoT) (sensors/devices)
- telecommunications (5G, nG)
- cloud
- artificial intelligence (AI)
- mobile applications
- WiFi-7
- Industry 5.0 [Javed et al. (2022)].

This foundation creates specific capabilities such as smart traffic monitoring, smart energy distribution, smart sewer systems, and many more. One, some, or all of these capabilities may be woven together to create the specific architecture for the given municipality. A connected community is not one thing; instead, it consists of a customized architecture of different emerging technology applications that are specific to the needs of the municipality. How the architecture is configured can have a significant impact on the citizens of the municipality (urban or rural), in addition to the critical infrastructure upon which the technologies are deployed.

3. DATA GENERATORS AND AGGREGATORS

With so many ways to think about a connected community, the best way is to think of it as a giant data generator and aggregator. In a 2021 study of published literature on smart cities, Ullah et. al. (2021) studied the top technological and organizational risks to connected community architectures based on appearances in peer reviewed articles. According to that study, the top two technological risks were IoT and big data integration, while the top two organizational risks were user data security and data safety. A given architecture might consist of tens of thousands of connected IoT devices. Those devices collect information constantly, possibly close to real time. All those devices are connected via a 5G, or ubiquitous WiFi connection, and they report their results likely to a cloud. In the cloud, some form of data analytics is performed, likely by AI.

The results of that analysis must be shown to human operators in some form, whether as near-real time data flow or as an analysis report. From there, some adjustment is made to the urban environment either automatically or by data-informed humans. As an example, placing connected IoT devices on the homes of people in a municipality to monitor their electrical use can have huge benefits for the grid and for the power generation plant serving the community. In this case, the IoT devices would be collecting real time electricity use data and transmitting it back for analysis. After the analysis is complete, municipal leaders may choose to change the electrical plant's output to mirror the demand more closely.

Whether we are talking about an electrical plant, sewer monitor, or traffic system, deploying tens of thousands of internet-connected devices in the municipal environment will result in enormous volumes of data being generated and aggregated. The economic and geopolitical value of data is hardly in doubt nor its ability to adversely impact individuals if not properly protected. A reality of connected community architectures, regardless of how they are configured, is that they will generate and aggregate huge volumes of data on both individuals and entire municipalities, potentially entire regions.

Bibri (2019) discusses the emergence of big data in the municipal environment, but from the perspective of contributions to sustainability and sustainable urban practices. The study does not, however, highlight the potential for exploitation of architectures by malicious actors nor the homeland security impacts of data aggregation at the municipal level. While it does discuss the need for public

privacy and security, this literature review was focused on the components functioning together as intended revealing a gap in security standards discussions in the connected community arena.

The U.S. government provided a specific standard in September 2020 for the "Security and privacy controls for information systems and organizations" in the National Institute of Standards and Technology (NIST) special publication 800-53 [NIST (2020)]. The document provides "a catalog of security and privacy controls for information systems and organizations to protect organizational operations and assets, individuals, other organizations, and the Nation from a diverse set of threats and risks, including hostile attacks, human errors, natural disasters, structural failures, foreign intelligence entities, and privacy risks." While 800-53 provides important practices and guidelines, it is necessarily high level and lacks the specificity demanded by a convergent system of different devices. Second, the standard, while used widely, is not compulsory, leaving connected community architectures in an uncertain state depending upon whether municipal leaders decide to demand adherence to the standard by policy or contract language. A system that displays the level of convergence seen in connected community architectures demands a more specific standard for both cybersecurity and privacy controls at the technical level and should be paired directly with municipal or state policy and assigned an accountable official.

4. PRIVACY AND INTEGRATED RISK AT THE MUNICIPAL LEVEL

Most discussions on the topic of online privacy surround an individual's right to security of data and agency of their personal data. This conversation is indeed important and the imperative to protect the data and maintain the privacy rights of individual users online is critical and should continue to be the subject of efforts to improve. The nature of connected community architectures is that they collect and aggregate the personal data from thousands or millions of individuals. Viewed through the lens of personal privacy, this issue requires significant attention as it presents an attractive target for would-be malicious cyber actors. The potential for criminal cyber activity, as well as state-sponsored, geopolitically motivated cyber actions, is extremely high and individuals should have some level of assurance on how their data is being collected, stored, and used. When viewed through the lens of the entire municipality, the collection of these data takes on a different characteristic.

The theft of the personally identifiable information (PII) of an individual or group of individuals through a cyberattack is a serious issue that deserves the resources of the proper authorities, and the best efforts of cybersecurity professionals, to prevent. Stepping back from the view of privacy as an individual issue, the larger, and perhaps more impactful issue, is around the privacy of the municipality. The exposure or theft of PII of an individual, with its public apologies and promises of free credit monitoring, is serious for the individual, but in nearly all cases would not rise to the level of a homeland or national security issue. In the case of a municipality of any size, the pooled data about the behaviors and working of that municipality, as collected by connected community architectures, represents a potentially frightening new aspect of privacy – the privacy of an entire municipality.

Spicer et. al. (2023) found a “sharp divergence between the smart city services being put in place by municipal administrators and the types of services residents want to see.” This finding raises questions about how aware citizens are about the individual data and privacy issues and the broader municipal scope of the issue. Architectures that are implemented should not only address direct issues with municipal functions but also include public education and communications plans to create an informed resident population.

Architectures provide data that help leaders analyze municipal functions and adjust to optimize for a given goal. For example, the reduction of traffic in certain areas at certain times or the distribution of electrical energy at peak and off-peak times. That same information provides insights that can just as easily be used for malicious purposes. In the transportation example, a municipal planner might use deployed IoT devices to measure what subway stations are the most crowded at what times to determine how many cars should be running at peak hours. That same data could be used by a malicious actor to determine the best area to place an explosive device for maximum impact. Similarly, efficient electrical energy distribution is key to ensuring equitable critical infrastructure services in growing urban environments. The same information could be used by a malicious cyber actor to determine the best grid(s) to disrupt with a cyberattack against the energy system.

Both examples above unambiguously represent homeland security threats that are far beyond the scope of the normal privacy policies and measures. An underappreciated and understudied aspect of installing a connected community architecture in any municipality is the potential for the collected and aggregated municipal data to become a significant homeland or national security threat. Privacy policies regarding connected communities should not focus only on individual privacy but on the privacy of the municipality. At the national level, the Department of Homeland Security (DHS), through the Cybersecurity and Infrastructure Security Agency (CISA), should study the risks to entire critical infrastructure sectors related to the number of connected community architectures in each region. A large city like New York, Chicago, or Los Angeles would clearly have potentially hundreds of thousands or millions of deployed IoT devices in their municipalities, making the risk more obvious than if a collection of small- or medium-sized municipalities had small architectures. Depending on where each was located and how they were configured, the risk to critical infrastructure from a theft of connected community data could be equivalent in either case.

The security of pooled data at the municipal level represents a potential homeland or national security issue if a malicious actor accessed the data and decided to use it as a whole, rather than to steal the PII of an individual or group of individuals. Policies and cybersecurity measures should be designed to account for the privacy of the entire municipality, leading to cyber incident response procedures that mitigate possible attacks against the broader community or region. The introduction of deployed IoT devices into our municipalities may be proven to be a necessity as we cope with growing urban populations, the need for higher agricultural yields, more efficient energy distribution, and more. However, by definition, these devices are connected or adjacent to critical infrastructure systems that were heretofore not connected to the internet. The introduction of tens or hundreds of thousands of potential access points where there used to be zero is a significant change in the risk profile for any critical service and it is made more important by the fact that these systems are serving some of our largest population centers. That makes for both a fertile ground for criminal theft of individual data and of potentially more dangerous theft of the municipality’s data. With the target this enticing and the impact this great, the first step towards more security in connected communities is through the creation of interoperability standards.

5. INTEROPERABILITY AND RESILIENCE

In October of 2022, a little-known industry group published a technical standard that most have likely never heard of. It was called Matter² and it was developed by the Connectivity Standards Alliance.³ What they created was a protocol standard that allows smart home IoT devices to work together regardless of brand. You may have a smart speaker built by Apple in your home and with Matter you can buy smart devices from Google and other companies and integrate them into your home network natively. The importance of interoperability can be overlooked but it is a critical element of cybersecurity, and it is particularly important for connected communities. Javed et al. (2022) found that interoperability was listed as the top requirement for future smart cities. The first major gap in the deployment of connected community architectures is in interoperability standards and there is a template for how to do it.

A search for connected community components will yield no shortage of companies that are happy to provide their solution to your municipality. As an example, one company (name omitted) will provide you with a package that includes:

1. IoT sensors in a variety of functions.
2. Private 5G network for connectivity.
3. Cloud infrastructure for data storage.
4. AI for data analytics.
5. A slick dashboard for monitoring all devices.

That is an end-to-end, turnkey solution that is attractive to municipal leaders who do not want to waste time and go through contracting processes more than once. The problem, easily visible to any cybersecurity professional worth their salt, is that this network is not resilient. A single vulnerability could potentially take the entire network down, since this is an end-to-end solution. Interoperability does not eliminate cyber vulnerabilities, but it does increase the potential that an attack will be stopped at one component in the chain. If all components are built by a single company, they are likely to have common vulnerabilities among them. If the architecture

includes components from a variety of vendors, it is less likely that a single vulnerability will bring down the entire system. This is called vendor diversity, and it is an excellent way to build resilience into any network of devices.

This was part of the reason behind the development of the Matter standard, as the alliance recognized the resilience inherent in this solution. If it was recognized for individual homes, how has it been overlooked for municipal environments? The imperative to create a protocol standard for interoperability is analogous to the discussion on individual privacy versus the privacy of a municipality. While it is certainly important to increase vendor diversity in home IoT, vendor diversity is extremely important for municipal IoT given its proximity to critical infrastructure. As more municipalities roll out plans for connected community architectures, they need to have the option to include interoperable equipment as a cybersecurity and resilience measure.

6. DEPLOYMENT STANDARDS

The next gap in deployments is the lack of minimum requirements for architecture deployments. Part of the attraction of the connected community concept is that it is not a one-size-fits-all solution that may or may not work for a given municipality. Communities can, in theory, choose for themselves which challenges they can solve using technology deployments and how to best roll them out for the community's needs. That flexibility should remain a feature of connected community deployments, but it is too important to leave entirely to the discretion of community officials. Connected community architectures have the potential to directly impact critical infrastructure, large numbers of citizens, and to devolve into actual homeland or national security issues. These realities demand the creation of minimum cybersecurity standards that apply to municipal environments. The National Institute for Standards and Technology (NIST) is well equipped to create such a standard through its Global Community Technology Challenge.⁴ With the help of CISA's Infrastructure Security Division,⁵ the federal government could create the minimum standard required to ensure a cybersecurity baseline for all connected community deployments.

² <http://tinyurl.com/23vhwcr>

³ <http://tinyurl.com/4km4zuat>

⁴ <http://tinyurl.com/yqyr2n66>

⁵ <http://tinyurl.com/mks6269m>

7. POLICY GAPS

The final gap is in the policy apparatus of municipalities. It is critically important that connected community architectures be chosen according to defined municipal challenges and aligned with a strategic vision. Municipalities should have accountable officials in place to oversee not only the deployment but the long-term operation of the system. Small issues like missing a firmware update on a single deployed sensor could result in an attack vector that causes extreme damage, and someone must be accountable to ensure the integrity of the system. The following recommendations are provided to help community leaders build the required foundation for successful connected community deployments.

1. **Unifying strategy:** a 2023 study of twelve cities in Spain with a total of 1,625 smart initiatives found that formal strategic planning was the main tool used in successful implementation of smart initiatives [Bolívar et al. (2023)]. Strategic guidance provides the vision for a connected community project and provides answers to questions about why certain decisions were made. A unifying strategy should give municipal officials, at any level and in any department, a piece of paper to which they can point to justify the actions they are taking. The strategy should be public in an effort to maximize transparency. Examples of issues that should be covered:

- overarching priorities
- specific problems to be solved
- potential challenges
- statement on risk identification and mitigation
- public outreach plan.

2. **Accountability trinity:** accountability is the key to ensuring that policies are carried out into action. In the municipal environment, there are three offices that must be filled with an individual who is individually accountable and not wearing multiple hats. Given the amount of data being generated, the privacy implications, and the potential for security risks, the following positions are critical for creating an accountability trinity that will ensure the operationalization of priorities from the “unifying strategy”:

- Chief Information Officer
- Chief Privacy Officer
- Chief Information Security Officer.

3. **Map of deployed devices:** one of the biggest threats to connected community architectures is a cyber vulnerability in a single, seemingly unimportant, deployed sensor. If that sensor does not receive, or successfully install, a critical firmware update or patch, the entire architecture is in jeopardy and the risk to critical infrastructure services increases. To ensure the integrity of the entire system, a live map of the real time status of deployed sensors will provide human operators with the ability to see potential issues and respond to them quickly. In the absence of such a capability, a single sensor could provide the access point required by malicious cyber actors, which is the first step in a downstream attack that could escalate to create effects exponentially more damaging than accessing a single sensor.

4. **Contracting language:** one of the most powerful tools municipal leaders have is their contracting language. If contracts stipulate that the vendor adhere to a set of standards aligned with the unifying strategy, vendors will have to adjust if they want the contract. Municipal leaders should deep dive into procurement processes and contracting language and find ways to ensure the security measures they prioritized in their strategy. This also gives the public the peace of mind to know that the strategy is not just words.

5. **Public communications plan:** transparency is foundational to any connected community plan and should include a robust public communications plan. At minimum, this plan should consist of the following elements:

- **Early outreach:** in this phase, municipal leaders should engage the public on the challenges they see and how they believe technology can solve them.
- **Priorities:** the priorities, through the unifying strategy, should be public and promoted, not buried on a municipal website.
- **Crisis communications:** in the event of a cyber event, the municipality should be prepared to communicate with the public and provide updates on the state of the crisis.
- **Public education:** the municipality should build in outreach that provide education about what purposes technologies will serve and how they will benefit the community. These programs should include technical literacy courses, upskilling, basic cyber hygiene, and privacy rights.

8. CONCLUSION

Connected community architectures are already being deployed in the U.S. and around the world, and for good reasons. Growing urban populations and the need to make resource distribution more efficient and equitable are driving the implementation of technological solutions. The rollout of 5G was a major driver of the technological convergence in the municipal environment, providing the bandwidth to support thousands more deployed IoT devices. It is possible that large urban environments of the future will require connected community architectures to function, so it is critical that these deployments be executed in a way that inspires public trust and prioritizes security and resilience. Deployed IoT devices that monitor critical elements of municipal functions are able to gain impressive insights that help planners and officials create better communities. There are also some risks that have to be recognized, planned for, and mitigated to the best of our collective abilities. Below are a few important factors that need to be taken into consideration when considering a connected community deployment:

- **Can the identified problem be solved by a technology solution?** There have been suggestions that deployed IoT and the right AI algorithms can cure all municipal ills from traffic problems to social inequality. The reality is that the scope of what deployed IoT devices can solve is limited. These solutions, as they exist today, are best at finding efficiencies and optimizing services such as electricity, traffic, or water/sewage services. They are also very good at increasing access to information such as through public WiFi programs or municipal mobile applications that allow for better access to services. However, there is a limit and a connected community architecture, no matter how well designed, will not solve every problem. It is imperative that municipal leaders spend time on what the problem actually is, what its secondary impacts are, and whether it is feasible for a technological deployment to solve it.
- **Does the municipality have the internal resources to manage the architecture long-term?** As with any technology project, there is a lifespan and maintenance tail that has to be accounted for by municipal leadership. Even if there are specific provisions in the contract for the company to provide services, the municipality still must have people who can monitor and evaluate the

performance of the system and ensure its integrity. A community without the accountability trinity, or without sufficient staff to stay engaged with the architecture over its lifespan, is destined for trouble. Part of the evaluation on whether to support and implement a project should be a self-evaluation that looks at the community's capacity to operate the system in the absence of vendor support.

- **In what ways is public engagement built into the deployment?** This is a multi-phased issue that must start at conception and run through upkeep and potential crises. Key to this is education of the public on technology literacy and basic cyber hygiene. Implementation of architectures without public outreach and education will also encounter problems throughout the life of the system in the form of potential trust issues.

Connected community architectures are already in effect in multiple U.S. and global cities, but they lack a basic level of standardization that would allow security and resilience measures to be implemented to protect vulnerabilities to critical infrastructure. These localized systems, even if implemented in small municipalities, could become the critical cyber vulnerability that introduces risk to national critical functions and critical infrastructure sectors. That kind of systemic risk ultimately trickles down to specific systems and individual components but can escalate throughout the national structure. Direct cyber vulnerabilities to critical infrastructure are reason enough to enforce minimum standards, but the potential for a breach of municipal or regional data could result in even more catastrophic events. Between these two vulnerabilities, basic interoperability standards should be created and implemented, and basic security standards should also be enforced. This is not a call for regulation but for a recognition that the technology convergence that is providing us with the insights to optimize our municipalities also carries the potential to catastrophically disrupt it. Connected community technology is exciting and may prove critical to resource distribution and services in the coming years as urban populations grow. The interest in these architectures as a cyber target will also grow and it is incumbent on cyber professionals and policymakers to start mitigating risks now.

REFERENCES

- Bibri, S. E., 2019, "On the sustainability of smart and smarter cities in the era of big data; an interdisciplinary and transdisciplinary literature review," *Journal of Big Data*; Article 6:25, <http://tinyurl.com/tu2bkdp>
- Bolivar, M. P. R., L. A. Munoz, and C. A. Munoz, 2023, "Identifying patterns in smart initiatives' planning in smart cities. An empirical analysis in Spanish smart cities," *Technological Forecasting and Social Change* 196, <http://tinyurl.com/4bkpump>
- European Parliament, 2023, "European Union Artificial Intelligence Act," December 19, <http://tinyurl.com/3449ysr3>
- Ismagilova, E., L. Hughes, N. P. Rana, and Y. K. Dwivedi, 2020, "Security, privacy, and risks within smart cities: literature review and development of a smart city interaction framework," *Information Systems Frontiers* 24, 393-414
- Javed, A. R., F. Shahzad, S. ur Rehman, Y. Bin Zikria, I. Razzak, Z. Jalil, and G. Xu, 2022, "Future smart cities: requirements, emerging technologies, applications, challenges, and future aspects," *Cities* 129, <http://tinyurl.com/4nvc7uc>
- NIST, 2020, "Security and privacy controls for information systems and organizations," Department of Commerce, NIST SP 800-53, Revision 5; September, <http://tinyurl.com/2s4vpa5n>
- Spicer, Z., N. Goodman, and D. A. Wolfe, 2023, "How 'smart' are smart cities? Resident attitudes towards smart city design," *Cities*, Volume 141, <http://tinyurl.com/ydkfxbbf>
- Ullah, F., S. Qayyum, M. J. Thaheem, F. al-Turjman, and S. M. E. Sepasgozar, 2021, "Risk management in sustainable smart cities governance: a TOE framework," *Technological Forecasting and Social Change* 167, <http://tinyurl.com/25k72bk3>
- White House, 2023, "Executive Order on the safe, secure, and trustworthy development and use of artificial intelligence," Biden Administration, October 30, <http://tinyurl.com/2rbvbnap>

HUMAN/AI AUGMENTATION: THE NEED TO DEVELOP A NEW PEOPLE-CENTRIC FUNCTION TO FULLY BENEFIT FROM AI

MAURIZIO MARCON | Strategy Lead, Analytics and AI Products, Group Data and Intelligence, UniCredit

ABSTRACT

The recent wave of enthusiasm for artificial intelligence (AI), accentuated by the advent of ChatGPT 3.5, has resulted in technology firms and businesses racing to harness the potential of increasingly sophisticated AI systems. Yet, the pivotal element for maximizing the benefits of these technologies, namely human engagement, is often overlooked. To navigate the complexities and opportunities of AI, companies must prioritize “human/AI augmentation” strategies. These strategies should focus on fostering AI awareness, education, and culture to empower employees with the knowledge to leverage AI effectively. Additionally, adopting innovative organizational change management approaches encourages AI experimentation, enabling the discovery of relevant use-cases. Crucially, pragmatic reasoning should try to reimagine the roles within an AI-empowered workforce, actively shaping the future instead of adopting a “wait and see” attitude. Establishing dedicated teams at the crossroads of AI’s potential and human considerations is essential. By implementing comprehensive, people-centric plans, organizations can unlock AI’s full potential, ensuring a harmonious integration that benefits not just the business but society at large. This holistic approach will pave the way for enhanced competitiveness and profitability in the AI-driven future.

1. INTRODUCTION

Artificial intelligence (AI) is undoubtedly one of the most important, if not the most, market trends of the day. Everyone is talking about it, addressing the topic from various perspectives, from technological to philosophical, with a huge number of articles, books, videos, documentaries, and products released in the past few months alone.

However, it is well-known that AI is not in fact a new topic at all. American computer scientist, John McCarthy, referred to “artificial intelligence” at the now-historic Dartmouth conference back in 1956, marking the beginning of AI as a standalone research area [Dartmouth (1956)].

The reason for the recent uproar can essentially be attributed to, if somewhat simplistically, the market introduction of ChatGPT 3.5 in November 2022 [OpenAI (2022)]. This was

the first time that anyone, through a simple registration of a free account, could directly test the potential of an advanced artificial intelligence system on what is most “human” in people: conversational interaction.

The impact on the public was so significant that ChatGPT became the fastest consumer application in history to reach 100 million active users [Hu (2023)].

It captured the attention of the top management of virtually every company in every sector, to the point that a recent report published by BCG (2023a) indicated that, by 2025, generative AI (GenAI) alone (i.e., ChatGPT and similar technologies) will cover 30% of the total AI market, estimated at around U.S.\$60 billion. This is truly astonishing considering that GenAI has only been in the news for about a year and a half, while AI has existed in some form for decades.

Consequently, referring to what is taking place as “hype” is not at all far-fetched, especially considering that algorithms and applications of “traditional” AI (i.e., non-GenAI) have been developed for years.

In the financial services sector, for example, models based on machine learning, a subset of AI, are already quite widespread. These tools allow financial institutions to effectively segment their customer base and accurately calibrate the corresponding risk profile, enabling high performance both in identifying the pool of customers who are most likely to repay the credit granted and in defining the best interest rate. Essentially, credit providers could, already through “pre-ChatGPT” AI, increase the volume of credit issued and the revenues generated from it while minimizing issues associated with customer insolvency. Apparently, such capabilities were not impressive enough to generate the same level of interest as there is in today’s AI, as driven by the recent chatbots based on large language models.

I personally began experimenting with ChatGPT from the first days of its public availability and, after the initial wave of natural enthusiasm, the risks and opportunities for individuals and societies at large became very evident, with GenAI holding the potential to considerably alter the world of work and already imposing itself at speed [Marcon (2023)].

This has led to a series of considerations focused on people, aimed at maximizing the potential benefits of AI, both for business productivity and, of course, for workers.

The reasoning derives, among other things, from three key elements:

- Despite the existence of numerous reports suggesting that AI could have significant implications on the job market, including potentially leading to massive job cuts,¹ I am not convinced that this would take place as rapidly as many believe. This is because it would lead to an increase in unemployment in countries that are unable to manage through existing welfare tools, resulting in economic, and potentially political, instability. Governments will not allow this to happen to the extent that many believe. It is also hoped that the increasingly important corporate theme of ESG (environmental, social, and corporate governance), especially the “S”, will play a role in substantially mitigating this risk in the short term.

- It is clear that without AI algorithms or applications, the discussion of their use would be redundant, since the very subject of discussion would be missing. However, it is equally true that the vast majority of the value that companies will be able to generate from AI will be derived from how their employees are able to adopt it positively and fully utilize it, possibly even reinventing their own way of working.
- While companies in information technology typically have personnel who “breathe” technology daily, this does not necessarily apply to other sectors, such as manufacturing or financial services, simply because it is not their core business. Thus, even in the digital/IT departments of non-tech companies, the understanding of what AI is, the risks it exposes, or the opportunities it offers, is not always high. And this is a factor that can limit the benefits that are expected to accrue from AI, in some cases quite significantly.

Based on the aforementioned considerations, it appears evident that companies need to start thinking in a structured way about a coherent architecture of “soft” AI initiatives. Not technological, but rather focused on people, with the intention of maximizing the benefits obtainable from new technologies in a sustainable way. In other words, companies may need to build functions that we could define as “human/AI augmentation”, focused on aspects such as:

- AI culture, awareness, and education
- AI-related experimentation through structured (organizational) change management approaches
- People impacts and roles re-definition.

The level of maturity of these three elements is not yet optimal and is evolving in the market. However, it is possible to discuss each of them, understanding their key elements, through which concrete actions can then be defined in individual corporate realities.

2. AI CULTURE, AWARENESS, AND EDUCATION

Viewing AI solely as a technological tool would be reductive, limiting AI to specific use-cases and failing to generate the momentum necessary for companies to truly transform around AI.

¹ Briggs and Kodnani (2023) found that there are “300 million full-time jobs potentially exposed to automation.” Daugherty et al. (2023) found that “40% of all working hours can be impacted by large language models (LLMs) like GPT-4.”

Looking at the numerous articles on the subject, it seems that there is at least one aspect that the market is not completely addressing, namely how to weave AI into the fabric of human intelligences that comprise a company. In other words, alongside technological programs, there must also be initiatives to develop a genuine AI culture.

This is where things get complicated because, in the first instance, what does it even mean to have an AI culture program?

To answer this question, we can generalize the concept of culture with reference to corporate culture, which, as is well-known, is a set of values, beliefs, and attitudes instrumental to a company in achieving its business objectives. This is crucial: corporate culture is not an end in itself, but a means to an end. Consequently, defining an AI culture program begins with asking: what do we want to achieve through AI in our company? Once this question is answered, it is then possible to determine how personnel should behave to achieve these goals and plan accordingly how to influence their behavior to change or complement existing practices (i.e., the culture).

There is a significant risk here, however. It is all too easy to employ AI to reap immediate benefits through process efficiencies, which, in the real world, typically results in increased automation and staff reduction (as anticipated above). This may justify the current lack of focus on a healthy AI culture: if people are replaced by automation, the element required for cultural change (i.e., the people themselves) is missing. The market currently rewards this approach.

Besides cost-cutting, the other potential benefit of AI is revenue increase, which is a much more complex issue and difficult to leverage. While it is easy to understand the efficiencies that can be gained by observing, for example, software auto-generation tools, it is much harder to predict how much revenues can increase through hyper-personalization of products, or how much more effective a marketing campaign would be if supported by GenAI. The result is that based on the information available, anyone choosing between certain efficiencies now or a potential increase in revenues in the future would opt for the former.

However, it is also clear that something must be done about this because, when considering the above alongside the rapid pace of technological progress and the sluggishness of regulatory and legislative bodies, the enormous risks of degradation of social infrastructure and welfare systems become apparent. This is where the topic of AI culture should come back into play.

Since companies do not exist in isolation but are embedded in society, to which they have fundamental responsibilities, it should be in their primary interest to encourage all employees to be proactive in innovation founded on AI. This innovation should aim to identify tangible opportunities for growth with measurable returns that can convince company leadership and markets to invest.

An AI culture program should, therefore, address the following: facilitate transparent dialogue between various corporate levels, making everyone openly aware that some roles will no longer be needed but that, thanks to the new tools available, many others can be created. Each individual should contribute pragmatically to create new, useful, and profitable products or initiatives, and improve what has always been done, without reservations in being assisted by a digital collaborator.

To achieve this, an AI culture program should consist of at least five elements:

- **Information (preparing the soil):** understanding the ever-changing context of AI, aimed at providing everyone with the necessary foundation to become familiar with it. This should be done in an informative and accessible manner to reach the largest possible audience: AI is, and will be, too important in our lives to remain ignorant of it.
- **Education (planting the seeds):** more specialized training to develop and learn how to use the new technologies that are rapidly emerging. Mastery of these tools supports the generation of well-founded ideas for new use-cases.
- **Innovation groups and brainstorming (sprouting ideas):** regular brainstorming sessions, in which solutions are sought based on clearly identified problems, or based on opportunities enabled by new tools or case studies. These sessions must also include the clear qualification of costs and benefits for realization: ultimately, solid ideas and accurate cost-benefit analyses are necessary to convince an investor to allocate capital.
- **Communication and recognition (harvesting and selling the intellectual fruits):** regularly inform the entire company of the progress being made and reward the most innovative and winning ideas, as well as virtuous behaviors. This generates healthy internal competition and the possibility of reusing what colleagues have built in other areas of the organization for their own function, scaling up AI faster.

- **Feedback and adaptation (improve the farming practices):** an AI culture program with the aforementioned characteristics has a decidedly bottom-up structure because it must be perceived as necessary primarily by the employees. For this reason, it is essential to regularly capture their impressions and make necessary adjustments to make it even more engaging: always keep in mind that the most brilliant ideas are born through positive employee participation.

In addition to all this, there should also be structured plans for career development and transformation of organizational processes centered on AI, adapting to what the company is achieving and the directions it is taking with everyone's contributions.

Companies' openness to the constructive use of AI will have a positive impact on their value in the medium term precisely because they will have demonstrated a tangible commitment to maintaining a healthy society. Conversely, those that have harmed it through the use of AI, aimed solely at cost-cutting (and hence staff reduction), will be heavily penalized, just as companies that do not pay attention to environmental impacts or do not act in the interest of communities are penalized today.

Structuring an AI culture program that aims to achieve the above objectives is, therefore, a win-win-win move for companies, employees, and society. It is worth taking action now to capitalize on the benefits that will undoubtedly result from it.

3. AI-RELATED EXPERIMENTATION THROUGH STRUCTURED CHANGE MANAGEMENT APPROACHES

When a new technology is available, new ways for releasing it and ensuring its adoption are necessary. Gartner (2023) reports that 45% of executives interviewed credited the hype around ChatGPT for the reason for increasing their AI investments. Additionally, the same research shows that 70% of organizations are currently exploring use-cases for GenAI, and 20% have already developed applications that are in pilot phases or have been rolled out in production environments already.

“
Companies' openness to the constructive use of AI will have a positive impact on their value in the medium term.”

As is often the case in similar situations, any plans to introduce a new technology at speed and at scale must take into account the technical feasibility of such a demand, as well as, even more simply, the number of technical professionals that the company has (i.e., the AI teams' capacity). It is extremely easy, in such a context, to run the risk of doing too much too soon, trying to generate many ideas and expecting that they can all be realized almost immediately (and, usually, at no cost).

Typically, in these cases a large number of working groups are established to identify use-cases in which AI can provide support. These take in contributions from many people and many areas of the company and, in each case, the participation of specialist AI staff is necessary to ensure that the discussions are practical from a technological point of view. This is a classic “bottom-up” approach to innovation, which, while admirable for valuing everyone's contributions, poses several challenges:

- “Bottom-up” work often results in very specific ideas with limited potential benefits, mainly because participants in brainstorming sessions usually have a view confined to their area only.
- Qualifying the business case for these use-cases is very labor-intensive, requiring the support of the limited number of AI experts who are usually already very busy with other tasks, causing bottlenecks and delays.
- The disillusionment of people, as only a fraction of the submitted use-cases get approved for development, directly driven by the previous two points, resulting in only a limited number of impactful ideas and higher-than-usual workload that leads to longer waiting times.

In my opinion, the most significant structural issue is that identifying very focused AI-based use-cases risks reducing AI to just another technology, rather than leveraging it as a transformative business factor.

Obviously, none of this is to say that specific use-cases should be discouraged. Many have been developed pre-ChatGPT and have certainly yielded excellent results, such as process automation and manual activity reduction. My message is simple: excessive reliance on traditional approaches when looking for innovative transformation may not be the best option, as it risks disappointment and unmet expectations.

Leveraging traditional approaches also stems from a fundamental misunderstanding. When the various departments of a company initiate brainstorming sessions to generate AI-based ideas, they are often actually referring to GenAI, which is transformative by its nature. Hence, it requires innovative approaches to maximize its sustainable and lasting value, mitigating the risk of seeing enthusiasm deflate and the bubble burst in the short run [Kestenbaum (2023), McKinsey (2023)].

While a “bottom-up” approach exposes companies to these issues, a purely “top-down” approach has its own drawbacks as well, since it neglects the perspectives of those closest to operational processes and value creation (e.g., exposure to, and understanding of, the end customers).

That is not all. GenAI systems, such as ChatGPT (or Google’s Bard, Anthropic’s Claude, or even Microsoft Copilot now), are not tools that perform a specific and limited task, they can rather be considered as advanced, and (almost) general-purpose personal assistants. They have the advantage that people in a company do not have to invent them, they already exist. Rather, people need to learn how to use them to understand by direct experimentation how they can add value to the organization.

For these reasons, a “selective bottom-up” approach, where ideas are generated not through large brainstorming sessions but through daily use of these tools by specific individuals, could be more effective. This approach could be organized as follows:

1. **Identify taskforce participants:** select a limited number of people with specific qualities, such as being highly skilled and talented, open-minded, with a desire to experiment and innovative thinking, and who can influence and promote solutions, etc.
2. **Deliver dedicated training sessions for these participants:** deliver trainings focused firstly on “mindset” towards AI/GenAI, as it is crucial to convey that ChatGPT, or its peers, is not merely a substitute for human labor, but should be considered an advanced personal assistant, useful for highly valuable tasks like work review, problem-

solving, or coaching; and secondly on “technical usage”, as without proper training GenAI can yield disappointing results. Technical sessions on, say, prompt engineering are indeed required for extracting maximum utility and long lasting satisfaction.

3. **Experiment and reflect:** allow time for these key people to experiment with ChatGPT, and the like, to understand its actual utility and how it could add widespread value. This could range from basic supporting tasks, like drafting documents, to profound applications like rethinking professional roles and processes in the company.
4. **Define benefits and set objectives:** after an agreed-upon period (e.g., three months), ask participants to outline ways to utilize these tools to improve performance, specifying the expected benefits and timelines for achieving them. The objectives should then be discussed with the managers, approved, communicated, and monitored.

In addition to the above, two useful enhancements could be made:

- Sustaining a community for exchanging ideas and achieved results, such as asking for support to accelerate the realization of the benefits and enable “cross-fertilization”, so that ideas from one area could be shared and reused in another.
- Regularly repeating this entire process with newly identified people, such as those based in other areas of the organization, to gradually expand the business functions participating in the change. A “train the trainer” approach may also be considered for faster scaling.

Adopting GenAI systems in this manner would offer several advantages:

- It could be executed in full compatibility with the “more traditional” approach of demand management for specific AI use-cases.
- It would contain the costs of using ChatGPT (or equivalent), as access would be granted only to selected people chosen for involvement in the process outlined.
- Users would master a “general-purpose” tool that they could apply to their area of expertise, where they are presumed to have maximum competence and, therefore, the ability to identify where and how benefits could be realized.
- AI technical teams would no longer be a bottleneck, as the use-cases definition would essentially be delegated to the staff selected for such experimentation.

- Individuals responsible for executing the business case that supports the identified use-cases would be motivated to achieve the objectives, as this would be seen as an opportunity for higher visibility.
- It would ensure value-added utilization of the tool (e.g., ChatGPT, Microsoft Copilot), mitigating the risk of its discontinuation due to suboptimal use.
- Periodically proposing the process to additional stakeholders would refine the training technique, allow for more ideas to be shared (e.g., from previous sessions), progressively engaging all areas of the company in a granular way, and delivering value exponentially.

Lastly, if we are truly committed to a sophisticated and comprehensive approach, we should take steps in parallel to avoid potential issues stemming from a perception of “elitism” by those not involved in the taskforce. Specifically, we should:

- Provide all company employees with access to ChatGPT (or an equivalent tool – and with all the necessary protections; for example, to prevent data leaks), albeit in a more basic or “downgraded” form (for free of usage costs, if possible), as a standard productivity tool. This would prevent disillusionment among those not selected in the process.
- Offer all employees access to a structured educational program on AI (as discussed above), ensuring everyone has a basic understanding of the subject with no discrimination, and promoting a widespread positive attitude towards AI.

This proposal is not trivial to implement, but it can certainly solve the challenges that are inherent in other, more traditional approaches. However, it has the disadvantage of not being able to predict the extent of the benefits that can be achieved in advance, which could lead to limited managerial commitment in the early stages.

On the other hand, the execution cost is low by design, given the limitation of access to GenAI systems to a restricted number of participants.

The critical success factor is undoubtedly the correct identification of the people to involve in this taskforce. These individuals, thanks to their skills, creativity, and pragmatism, can maximize the chances of finding the desired value, which, once found, will become the engine of subsequent iterations.

If this were to prove true, it would once again demonstrate that AI's success ultimately depends on human intelligence.

4. PEOPLE IMPACTS AND ROLES RE-DEFINITION

In a recent study published by BCG, it was shown that only a small fraction of company employees (14%) have received training courses explaining how their jobs would change as a consequence of the advent of AI, even though a majority of them (86%) feel the need for such knowledge [BCG (2023b)]. Furthermore, various articles and interviews have shown that when executives are typically asked “how do you expect the roles of employees in the company to change due to AI?”, their answers are almost always vaguely along the lines of “I am very curious to see how the roles will change.” This simply indicates that there is still a lot of uncertainty on the subject.

It is clear that there are dozens, sometimes hundreds, of roles within a company, making it difficult to provide concise answers. However, in my opinion, it is possible to deduce the broad impacts starting from a mid-level position in the hierarchical pyramid: the middle management. From there, one can begin to extend the implications both upwards and downwards within the pyramid and envisage, as a target, what roles might look like in the medium term for companies fully supported by AI.

This reasoning is based on a fairly recent personal experience.

A while back, I found myself proposing a survey to measure the morale of our team members subsequent to an organizational change, which typically has an impact on the mood and motivation of people. After going through the necessary steps and receiving the green light to proceed, a junior colleague and I began to work on it. Neither of us had ever conducted such a survey before, but with ChatGPT available in the market, we certainly had a powerful tool at our disposal to support us.

We basically had two options. We could either ask the chatbot to prepare the questionnaire for us, and then submit it to the manager and HR colleagues for review, or reflect independently on how the survey should be constructed, prepare a first version with our thoughts, and only then use ChatGPT as a reviewer.

Clearly, the first solution, despite being quicker and requiring less effort, could significantly diminish our “corporate utility”. Letting ChatGPT do the thinking for us, which I guess is not the best of ideas, could also lead to unsatisfactory results. Consequently, we decided to go for the second option.



We first reflected on four or five macro areas that would have been reasonable to address to probe the team's morale, and then dedicated ourselves to defining a limited series of questions for each area that would bring out what we were looking for, coming up with about 20 questions in total. At this point, we submitted them to ChatGPT, providing the appropriate context, and asking for both an evaluation and possibly suggestions for improvement.

The chat gave us a rating of 7/10, which, as we were aiming to do a good job, we did not consider to be sufficient. We then refined the questions, also incorporating the recommendations provided by the AI in the first iteration, and resubmitted them. Our rating improved to 8.5/10. At this point, after no more than two hours of work and reflection, we were satisfied and shared them with the team manager and the HR contact, who both approved the survey with no changes.

In just two hours, despite having no prior knowledge of the subject and relying on both our human intelligence and artificial support, we prepared a piece of work that was far from trivial. Typically, such work would in fact require an expert in the field, yet ours needed no modifications.

This struck me so profoundly that I began to ask myself questions that in practice all converged towards the following: if in the near future, company staff have access to tools that allow them to produce high-quality work that managers no longer need to review, how would their role have to change?

This is still a somewhat hypothetical question, as I am assuming that the entire staff of a company reaches a level of maturity that makes this scenario plausible. However, the thought experiment is useful for outlining ideas.

A first response to the above example, more instinctive than rational, is to think that "managers are no longer needed", but this would clearly be an oversimplification. It is true that AI will impact, and in some cases replace, the work of not only operational staff but also the "white collars".² But it is equally true that new AI technologies will bring about radical changes in scenarios compared to the present, which will necessitate significant evolutions, rather than the elimination, of managerial and administrative roles.

² The CEO of IBM recently stated that "I actually believe that the first set of roles that will get impacted are [...] white-collar workers" [Chiang (2023)].

To draft the future role of a manager in this context, we must start with an almost obvious assumption: managers will see a significant reduction in their traditional roles as taskmasters and supervisors. This is because people, equipped with all the necessary (AI) tools, will become much more efficient. They will produce work faster and of higher quality, reducing the need for a traditional managerial figure. This shift will undoubtedly be a difficult adjustment for many managers. They will have to come to terms with not being perceived as the most competent and skilled members of their team, as they once were.

This change represents a profound revolution in the manager's role, which might lead some to resist this new reality.

To overcome such resistance, adopting a mature approach to this epochal transformation is essential, and new paradigms can already be imagined to turn the threat (for the manager) into an opportunity (for everyone). In particular:

- **Shift from managing people to partnering with them:** many managers like to proudly refer to “my people”, indicating a hierarchical relationship. While this is normal, in a scenario where team members can independently produce excellent work, maintaining a stance as the most competent person, perhaps finding instrumentally the famous needle in a haystack, and to continue to impose one's organizational authority could lead to staff frustration. An approach where a manager acts more as a partner is, therefore, more suitable. This is because people will always need to discuss their work with someone they respect. The focus, however, is increasingly shifting towards higher-level concepts, ideas, and paths for collaborative growth, rather than on the quality of the deliverables. Consequently, being an authoritative figure who listens attentively and engages in meaningful dialogue will be highly valued.
- **Acceptance of sharing “their” people with other managers:** this is a natural extension of the previous point. As individuals develop their skills and exploit AI tools, they may generate original and innovative ideas that transcend their current area's scope, potentially benefiting other areas as well. Far from being detrimental, cross-team and cross-level collaborations are highly beneficial for the company, fostering value, creativity, and innovation.

Consequently, managers need to move away from a strictly controlling attitude, which includes limiting team members' interactions within the organization. Instead, adopting a smart “open (re)source” style of management, which encourages expanding one's professional network and facilitating connections, is a more effective and forward-thinking approach.

While the two points above indicate a shift towards more fluid and open relationships between staff and managers, without a proper “glue” this shift could clearly lead to organizational chaos.

As a consequence, for managers to succeed, their focus should increasingly shift from overseeing tasks to building binding elements that maintain team order and stimulate a proactive willingness among team members, such as:

- Nurturing a mission that inspires team members to naturally contribute and add substantial value.
- Acting as an advocate for their area within the company, promoting it beyond traditional boundaries and established networks, thereby extending its influence and making it a magnet for ambitious talents.
- Being a proactive agent of innovation, continuously refining the organizational structure with both their own and their team members' ideas to foster growth and expansion.
- Recognizing deserving team members transparently and honestly by, for example, granting them more visibility and autonomy, and ensuring they feel part of the broader organizational objectives and mandates.

These strategies, while not entirely new, gain fresh relevance when executed in conjunction with the aforementioned partnership approach and fostering cross-functional relationships. This combination reshapes the managerial role, aligning it more closely with that of an entrepreneur and networker.

It also means that managers will not entirely “own” their teams. Instead, they will share “stakes” in them with team members who contribute the most in broadening the scope of their functions and enhancing their relevance within the broader organization.

At this point, another key element becomes clear: at least some of these new characteristics of the manager are already present today in the mandates of top-level executives. Consequently, if we accept the flow of the discussion so far, then the entrepreneurial expectation could extend to lower managerial levels, leading to two significant outcomes:

- **Simplification of the organizational structure:** by shifting entrepreneurial tasks to lower managerial levels, the organization might reduce its hierarchical layers, as some of these may become redundant.
- **Fostering of innovation and the possibility of creating new areas within the organization:** increasing freedom to act within the company, coupled with a greater focus on creativity and innovation by each team member, can potentially lead to such a broadening of the team's activities that new, self-contained areas may be created.

If the outlined reasoning is sound, then a manager who resists the initial elements of partnership and “sharing” their team members with other managers risks creating a static environment or, worse, widespread team frustration. Such resistance could lead to a scenario where the manager's role really becomes redundant.

On the other hand, if a manager embraces these changes and evolves towards a more entrepreneurial and networker approach, they not only solidify their leadership but also become a driving force for organizational expansion through innovation.

And it is this evolved stance that could lead to the development of new, independent managerial roles, possibly through the spin-offs of existing departments that have charted a significant new direction.

5. CONCLUSION

Talking with many people about AI, I often get a sense of uncertainty and even fear of the future. This stems not just from the deliberately exaggerated, and sometimes polarized, communication by the media, but also from the fact that we are truly at the beginning of an era that it is entering our lives with phenomenal speed.

However, since companies are a fundamental element of society, much of the responsibility for creating a healthy and sustainable world falls on them. By introducing organizational structures that have the mandate to support people in constructively adopting AI, they can also help make this transition smoother and less painful.

The three pillars outlined above (i.e., AI awareness, education, and culture; AI experimentation through innovative – organizational – change management approaches; and people impacts and roles re-definition) thus represent both a conceptual and an operational framework for structuring the foundations of a function that deals with an actual human/AI augmentation. With these, a company can derive the actions it deems necessary, depending on its own mandate and priorities.

What will make building and working in such an area exciting and truly rewarding is its mission: to contribute to delivering true, positive, and sustainable value to people, companies and, therefore, society at large.

REFERENCES

- BCG, 2023a, "Generative AI," Boston Consulting Group, <http://tinyurl.com/bdzhc96k>
- BCG, 2023b, "AI at work: what people are saying," Boston Consulting Group survey, June 7, <http://tinyurl.com/2mbk2krh>
- Briggs, J., and D. Kodnani, 2023, "Generative AI could raise global GDP by 7%," Goldman Sachs, <http://tinyurl.com/mtsdu3ku>
- Chiang, S., 2023, "IBM CEO says AI will impact white-collar jobs first, but could help workers instead of displacing them," CNBC, August 22, <http://tinyurl.com/2t54afcc>
- Dartmouth, 1956, "Artificial intelligence coined at Dartmouth," The Dartmouth Summer Research Project on Artificial Intelligence was a seminal event for artificial intelligence as a field, <http://tinyurl.com/25sje5um>
- Daugherty, P., B. Ghosh, K. Narain, L. Guan, and J. Wilson, 2023, "AI for everyone," Accenture, March 22, <http://tinyurl.com/25fdjchv>
- Gartner, 2023, "Gartner poll finds 45% of executives say ChatGPT has prompted an increase in AI investment," press release, May 3, <http://tinyurl.com/bdeyutrx>
- Hu, K., 2023, "ChatGPT sets record for fastest-growing user base – analyst note," Reuters, February 2, <http://tinyurl.com/284zdzz2>
- Kestenbaum, R., 2023, "ChatGPT is losing users. Is the artificial intelligence craze over?" Forbes, July 11, <http://tinyurl.com/445j3j2c>
- Marcon, M., 2023, "The ChatGPT phenomenon: people must remain in the driving seat," LinkedIn Insights from the community, January 17, <http://tinyurl.com/5dhanf2a>
- McKinsey, 2023, "What's the future of generative AI? An early view in 15 charts," McKinsey & Co., August 25, <http://tinyurl.com/yc59yh9n>
- OpenAI, 2022, "Introducing ChatGPT," blog, November 30, <http://tinyurl.com/4frc3vhn>

BUILDING FINTECH AND INNOVATION ECOSYSTEMS

ROSS P. BUCKLEY | Australian Research Council Laureate Fellow and Scientia Professor, Faculty of Law and Justice, UNSW Sydney¹

DOUGLAS W. ARNER | Kerry Holdings Professor in Law and Associate Director, HKU-Standard Chartered FinTech Academy, University of Hong Kong

DIRK A. ZETZSCHE | ADA Chair in Financial Law, University of Luxembourg

LUCIEN J. VAN ROMBURG | Postdoctoral Research Fellow, UNSW Sydney

ABSTRACT

In our new book, *FinTech: finance, technology, and regulation* [Buckley et al. (2024)], based on analysis of experiences with the integration of new technologies into finance and of digital financial transformation around the world, we present strategies for policymakers and regulators seeking to build FinTech and innovation ecosystems, to support digital financial transformation and inclusive, resilient, and sustainable digital finance. This strategy comprises three levels. First, we focus on the central role of digital public infrastructure and digital financial infrastructure, based on a four-pillar strategy, which includes: (i) digital ID and e-KYC systems; (ii) open, interoperable electronic payment systems; (iii) electronic government provision of services; and (iv) enabling new activities, business, and wider development. Second, we set out seven elements that encompass appropriate regulatory approaches to support digital finance. Finally, we highlight the role of the wider ecosystem, focusing in particular on data strategies and support for research and innovation. This strategy is central to balancing the risks and opportunities of digital finance, FinTech, and innovation to contribute meaningfully to the advancement of inclusive sustainable development.

1. INTRODUCTION

In the next decade, global finance will be significantly impacted by the rapid advancement of technology, the need for sustainable development, and the perennial friction between economic, financial, and technological fragmentation and globalization. While these developments may introduce novel opportunities, they may also present challenges for the financial system.² Digital finance, if correctly designed and regulated, can be applied to ameliorate the effects of future crises and can advance inclusive sustainable development. Regulators, who form an integral part of the interactive system that financial technology (FinTech) encompasses, are required to implement strategies to ensure that the financial system is fit for the future.

On this basis, this article seeks to analyze which strategies regulators are required to implement to ensure that the financial system is fit-for-purpose going forward. These strategies have been formulated based on a synthesis of core lessons drawn from the past decades, and focus on building the necessary digital infrastructure and developing new regulatory approaches.

2. DIGITAL FINANCIAL INFRASTRUCTURE

Digital financial infrastructure is central to advancing the aims of finance, from financial inclusion to financial stability, resilience, and sustainability. The experiences drawn from various crises have reinforced the significant role of digital

¹ We would like to thank the ARC Laureate Fellowship Scheme, the Hong Kong RGC Senior Research Fellow Scheme, and the ADA Chair in Financial Law (inclusive finance) for financial support.

² Buckley et al. (2024) sets out the factors that will impact this outcome.

infrastructure as being fundamental for crisis management, economic recovery, and sustainable development. In our view, countries need to direct their focus to four pillars of digital financial infrastructure (set out below), which are essential to supporting digital financial transformation [Arner et al. (2021)]. The adoption of such a strategy may realize the full potential of FinTech on the basis of a progressive approach to the development of the underlying infrastructure for digital financial transformation.

2.1 Pillar I: Digital ID and e-KYC systems – establishing the foundation

Mobile payments and the required foundational layer of digital identity (digital ID), specifically sovereign digital ID, are central to the digital transformation process, and constitute the required foundation for all subsequent components of a digital financial ecosystem. Several digital ID systems have been developed, particularly to assist less advanced economies, where people lack formal identification documents. IrisGuard, for example, is a digital ID solution composed of iris recognition technology that converts an iris image into a unique code, which is subsequently used to identify an individual. IrisGuard has developed digital ID solutions for the U.N. and Jordan that focus on digital ID solutions for refugees.³

IrisGuard's digital ID solutions provide the necessary digital ID to enable beneficiaries to receive food vouchers, withdraw cash, and to transfer funds without the need for a bank account. To this end, it provides what we refer to as the "base ID infrastructure", which establishes a link between the physical individual and the specific digital service. While IrisGuard's digital ID solutions make use of human physical attributes, base ID can also be developed from several sources, which include business-specific electronic identities, such as customer accounts with e-commerce platforms. India's Aadhaar system is another example of base ID, which entails the issuance of a 12-digit randomized number to all Indian residents and facilitates access to financial accounts and digitizes government payments and services.⁴

Base ID, therefore, provides a fundamental element necessary for the "know your customer" (KYC) process. The central aim is to enable bank account opening for the majority of people and entities in a simple and cheap manner. This permits resources to be redirected towards the protection of market integrity and for analyzing the position of high-risk

customers. The technology thus enables the interlinkage of various systems, which assists with balancing economic growth, financial inclusion, and market integrity while complying with international financial standards. In Europe, for example, the eIDAS system interlinks the ID systems of the 27 E.U. member states and bank account opening without physical attendance.⁵

Digital ID systems can also be used to store customer financial criteria to enable financial institutions to identify customers' needs and preferences from a stronger starting point. Electronic identification is, therefore, required as the foundation from which financial institutions comply with customer due diligence standards, thus enabling a wider array of financial services. It should be noted, however, that while technically possible, the interconnection of digital ID systems may not always be politically feasible. Cybersecurity and data protection challenges may also thwart the unwavering support for mandatory, all-encompassing digital ID systems for all members of society.

2.2 Pillar II: Open, interoperable electronic payment systems – building connectivity

Access to payments must be ensured once a digital ID system has been developed. Payment systems establish the fundamental infrastructure through which money flows in any economy. One way in which FinTech can assist is through advancing a mobile money (e-money) ecosystem. In general, e-money is defined as a stored value instrument or product that: (i) is issued on receipt of funds; (ii) consists of electronically recorded value stored on a device such as a mobile phone; (iii) may be accepted as a means of payment by parties other than the issuer; and (iv) is convertible back into cash [Binda (2020)]. Mobile money enables the payment of bills, remittance of funds, and deposit of cash through the use of a mobile phone.

Interoperability is key to the impact of digital payments, which governments are mandating increasingly to expand economic and social benefits and innovation. Crucially, digital payments are made attractive through the use of interoperability to bring together traditional and new forms of payment. In China, for example, Alipay and WeChat have illustrated the power of the facilitation of new entrants and the digitization of traditional payment systems. By 2021, 64% of the Chinese population made use of mobile payments, with Alipay and WeChat

³ <https://tinyurl.com/2xzwf25v>

⁴ "About Aadhaar", Unique Identification Authority of India, <https://tinyurl.com/w55j5ypz>

⁵ "eIDAS: The Digital Identification Regulation for Europe," ElectronicID, <https://tinyurl.com/2ry3e9je>

constituting 91% of all digital payments effected [Schirmer (2022)]. Overall, governments are increasingly mandating interoperability as a licensing condition for payment providers.

2.3 Pillar III: Electronic government provision of services – expanding usage

The use of open electronic payments infrastructure by governments, as provided for in Pillars I and II above, is integral to the process of digital transformation, and can be effected through state support payments made through digital government-to-person (G2P) payments. These digital payments support governments in their shift from in-kind assistance (i.e., supply of food and water) to more affordable cash transfers. Further, accounts used for G2P support payments may also be used for non-government payments purposes. They also support financial education initiatives by enabling people to learn how to use digital payments through the relevant digital platform. G2P payments can, therefore, be used to further financial inclusion and sustainable development.

G2P payments have been used by several governments with the aim of bringing the financially excluded into the formal financial system and to enhance the efficiency and effectiveness of government payments, services, and transfers. During the COVID-19 pandemic, the use of such payments increased significantly, with 60 low- and middle-income countries making use of digital assets or payments to deliver social assistance programs [World Bank (2021)]. In Tunisia, the first round of emergency COVID-19 payments was delivered via the post. However, during the second round of payments, users were able to register for their payments digitally, in addition to selecting their preferred digital payment method.

G2P payment systems should, however, be properly designed to facilitate the achievement of the aforementioned objectives. In general, well-designed G2P payments comprise three fundamental characteristics. First, account procedures should later facilitate unrestricted payments. Second, the digital-to-real gap should be bridged as individuals will prefer cash where digital transaction partners are limited. If merchants are unable to conduct their business without the acceptance of e-money, experience illustrates that they will provide devices that accept e-money efficiently, with or without incentives. Finally, functionality should be simple and must enable learning for making and receiving transfers.

Governments can advance digital transformation by highlighting the advantages of using e-money, by requiring merchants to accept digital payments at low or no cost to customers, and through setting limits for cash transactions in the real economy. Overall, G2P payments may be used to facilitate improved tax collection, as small and medium enterprises advance within the formal financial system, in place of developing outside of it. It may also provide support for the development of national pension systems over time, which enhances the available financial safety net and the provision of additional financial resources to drive growth.

2.4 Pillar IV: Enabling new activities, business, and wider development

The digital infrastructure created in Pillars I – III can be built upon to create innovative forms of financial services that enable new activities and business, and broader development. For example, in place of the traditional provision of credit through credit risk analysis conducted by specialized banks on the basis of collateral, digitization has allowed for the pricing of credit through datafication (i.e., the process of using and analyzing data). More accurate data may, therefore, be gathered from e-commerce platforms, search engines, social media services, and telcos [Zetzsche et al. (2018)]. The big data approach of TechFins⁶ potentially enhances business decisions through the provision of a better picture of customers' financial positions using the more accurate datasets.

TechFins can, therefore, play a central role in re-personalizing the financial relationship with their customers through adjusting credit rates on the basis of the real risk profiles of individuals. Further, transaction costs per customer are significantly lowered on the basis of the economies of scale inherent to the tech platforms used by TechFins. As a result, the provision of "personalized" services at a reduced cost per customer supports the delivery of financial services for small amounts of money, which also advances financial inclusion. However, in spite of the potential significant benefits, the introduction of TechFins also creates novel challenges at the intersection of data and financial regulation.

The increased access to, and reduction of transactions costs for financial services provided by digitalization also advances the expansion of the level, range, and quality of insurance and investment services, and supports the progression of technologies such as artificial intelligence (AI).⁷ This expansion

⁶ TechFins have been described as established technological and e-commerce firms who provide financial services [Zetzsche et al. (2018)].

⁷ See Part II of Buckley et al. (2023)

and progression may possibly bring new financial services into the financial system that may correspondingly advance business development, infrastructure, and innovation, through increased savings rates, which may be redirected through the financial system.

3. NEW FINANCIAL REGULATORY APPROACHES

Digital infrastructure requires appropriate legal and regulatory frameworks that support the creation of inclusive, resilient, and sustainable digital finance. To ensure that the financial system is able to contribute to the achievement of these objectives, digitalization must be paired with fit-for-purpose regulation. In our view, governments and regulators alike should direct their attention to the implementation of new financial regulatory approaches that can act to support the achievement of the aforementioned objectives. We set out below these new regulatory approaches in seven principal elements.

First, a broader analytical framework is required to address the risks associated with innovation, including: (i) new sources of traditional risk, (ii) new forms of risk, and (iii) new markets and systems. The application of such a framework requires a careful consideration of the principal areas of concern that have emerged during the process of digital financial transformation. These areas of concern include cybersecurity, data security, and data privacy, and the appearance of new systemically important data-driven financial institutions, such as novel forms of market infrastructure.

Second, regulators should expand their expertise to continuously deepen their understanding of the interlinkages between the real economy and finance, which is growing ever more complex. Multidisciplinary insights are required, spanning the social sciences, and the formal natural sciences, to reassess and account for different risk exposures and to evaluate its impact on sustainability and the recovery of each regulation. Practically, this means that regulators must recruit more staff with analytical, interdisciplinary, and scientific skills, with expertise in, for example, the subject of systems science, to properly understand how climate change is likely to impact various environmental risks.

Third, regulators should promote innovation through the adoption of balanced proportional risk-based regulation. To ensure that the financial system is fit-for-purpose, they should assess and modernize ill-suited regulation as identified

“

Regulators will be required to make increasing use of technology to effectively regulate finance.

”

through regulatory impact assessments. Such assessments will assist in determining whether legacy rules remain helpful in the modern era of financial regulation. In parallel thereto, regulators should also put into place effective regulatory facilitation arrangements, such as innovation hubs and regulatory sandboxes, which promote innovation and mutual learning through extensive interaction between themselves and various market participants [Buckley et al. (2024)].

Fourth, regulators will be required to make increasing use of technology to effectively regulate finance and they should be greatly aided by higher levels of datafication. To this end, they should upgrade their supervisory data systems and regulatory technologies (SupTech and RegTech) and work towards supporting the developing of core digital infrastructure. The focused roll-out of SupTech and RegTech will support apt proportional regulation of innovative financial products and services, including those underlined by principles of sustainability, which will have the net effect of benefiting financial inclusion and will address risks associated therewith.

Fifth, regulators and regulated entities will likely need to adopt a “beta approach” to regulation, which is common to software development. Regulation will never be without fault and the use of finance to mitigate the effects of external shocks will require a continuous adjustment to the relevant rules and standards. As such, financial regulation will be informed by a process of trial and error, in which regulators learn from experience and adjust on the run, an approach that may likely be abhorred by many traditional regulators. This will likely require a combination of hard and soft law, in addition to binding and indicative forms of regulation.

Sixth, the efficacy of financial regulation and sustainable development can be further progressed through regional regulatory approaches that support the needed scale. To this end, regulators and policymakers in many countries

will be required to support regionally harmonized regulatory frameworks. Consistent regulatory approaches across a specific region will advance national markets' interests in innovative financial service providers. At the same time, the increased concentration of providers in a region will provide consumers with a wider range of services to choose from, while benefiting from more competitive prices. It would also increase the possibility of providers creating innovative solutions to a broader range of issues.

Finally, regulators will be required to consider the broader societal ecosystem in which FinTech operates to advance inclusion, innovation, resilience, and sustainable development on the basis of a much-widened regulatory mandate. In the context of digital finance, the broader ecosystem focuses on education, funding, and skills, in addition to the development of expertise on the basis of related professional and other associations. In many countries, greater focus has been directed towards advancing education in the STEM disciplines and in social science research into the effect of technology on humankind.

A focused implementation of these new financial regulatory approaches should likely encourage and facilitate the advancement of innovative financial products and services, while at the same time attending to the corresponding risks. They form an overarching strategy that supports FinTech, financial inclusion, innovation, and sustainable development, and can be further progressed through a requisite focus on building digital infrastructure and on regional regulatory approaches that support the required scale.

4. THE WIDER ECOSYSTEM: DATA, RESEARCH AND INNOVATION SUPPORT

The third level of the strategy involves the wider ecosystem. From the standpoint of the wider ecosystem, three elements are particularly important: an enabling legal system, strategies to maximize the benefits of data, and approaches to support research, development, and innovation.

From the standpoint of the legal system, it is important – in addition to regulatory approaches at the second level and infrastructure at the first – to consider the role of private law in providing appropriate support. This relates directly to the need for strategies to maximize the benefits of aggregate data while minimizing the risks of concentration and dominance, which result from combinations of network effects of technology and economies of scope and scale of finance. These include both clear legal approaches to data as well as frameworks to support sharing and use, particularly mandatory open finance. Finally, these are enabled by support for innovation through mechanisms such as innovation hubs and research and development funding. Together, this combination of an enabling legal system, including for data, along with strategies for maximizing the benefits of data along with support for innovation, research, and development, provides the wider context to support digital financial transformation and inclusive sustainable development.

5. CONCLUSION

The central aim of this paper was to set out strategies to be implemented by regulators for building digital financial infrastructure that supports digital financial transformation. At the same time, it aimed to set out new financial regulatory approaches that can be adopted to support inclusive, resilient, and sustainable digital finance. To ensure that the financial system is capable of contributing towards the achievement of these roles and objectives, digitalization is required to be paired with fit-for-purpose financial regulation.

These lessons can support governments and regulators in ensuring that digital finance is appropriately enabled, supported, and regulated in order to mitigate the effects of future crises and best contribute to the advancement of inclusive sustainable finance.

REFERENCES

Arner, D. W., R. P. Buckley, D. A. Zetsche, and R. Veidt, 2021, "Sustainability, FinTech and financial inclusion," *European Business Organization Law Review* 21:1, 7-35

Binda, J., 2020, "Cryptocurrencies: problems of the high-risk instrument definition," *Investment Management and Financial Innovations* 17:1, 227-241

Buckley, R. P., D. W. Arner, and D. A. Zetsche, 2023, *FinTech: finance, technology and regulation*, Cambridge University Press

Schirmer, A., 2022, "Payment methods in China: how China became a mobile-first nation," *Daxue Consulting*, <https://tinyurl.com/3n3xeyaz>

World Bank, 2021, "Identification for Development (ID4D) and G2Px, annual report 2021," <https://tinyurl.com/bm6jvh2w>

Zetsche, D. A., R. P. Buckley, D. W. Arner, and J. N. Barberis, 2018, "From FinTech to TechFin: the regulatory challenges of data-driven finance," *New York University Journal of Law and Business* 14:2, 393-446k, March 18, <http://tinyurl.com/99r6xmre>

USE AND MISUSE OF INTERPRETABILITY IN MACHINE LEARNING¹

BRIAN CLARK | Rensselaer Polytechnic Institute
MAJEED SIMAAN | Stevens Institute of Technology
AKHTAR SIDDIQUE | Office of the Comptroller of the Currency

ABSTRACT

Machine learning methods, the foundation of much of artificial intelligence (AI), are now widely used in data analysis and model-building across a variety of disciplines. These techniques have also become the underpinnings of many of the business intelligence (BI) analytics that are being widely deployed across a wide range of industries. In this article, we focus on some elements of inference around analytics possible in machine learning, contrasting them with how applied econometricians traditionally approached inference. We do this in the context of applying both traditional econometric methods and several machine learning methods to the same dataset.

1. INTRODUCTION

Machine learning methods, the foundation of much of artificial intelligence (AI), are now widely used in data analysis and model-building across a variety of disciplines. These techniques have also become the underpinnings of many of the business intelligence (BI) analytics that are being widely deployed across a wide range of industries.²

With freely available software such as Keras, Tensorflow from Google, lightGBM from Microsoft, and Torch from Facebook, the techniques have also become widely available. The provision of such open-source software, accompanied by the rise of cloud-based platforms from Amazon, Google, Microsoft, etc., have significantly reduced the need to build out hardware infrastructure. Historically, machine learning has focused much more on prediction than on statistical inference around analytics.

In finance, machine learning (ML) and deep learning (DL) have been applied extensively to credit risk modeling (e.g., default prediction) due to the availability of a large quantity of data. Butaru et al. (2016) have applied machine learning to credit cards. Sadhwani et al. (2021) have applied deep learning to mortgage risk. These have largely been focused on forecasting delinquency or defaults.

Traditionally, econometrics has also analyzed data and built models. In contrast to data scientists, econometricians have traditionally focused significantly on statistical inference. Biddle (2017) provides an overview of how statistical inference has changed over time. His definition of statistical inference – “the process of drawing conclusions from samples of statistical data about things that are not fully described or recorded in those samples” – describes what econometricians do fairly well.

¹ Views and opinions expressed are those of the authors and do not necessarily represent official positions or policy of the Office of the Comptroller of the Currency or the U.S. Department of the Treasury.

² Korolov, M., 2018, “New AI tools make BI smarter — and more useful,” CIO Magazine, April 18, <http://tinyurl.com/yyuszz9k>

Figure 1: Shapley values based on lightGBM

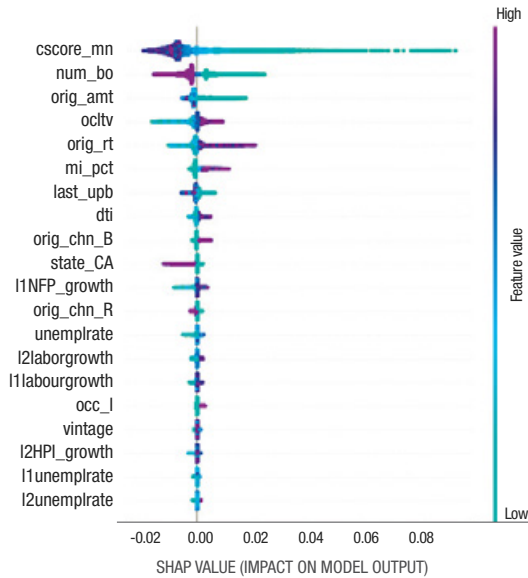


Figure 3: Shapley values based on deep learning/Keras

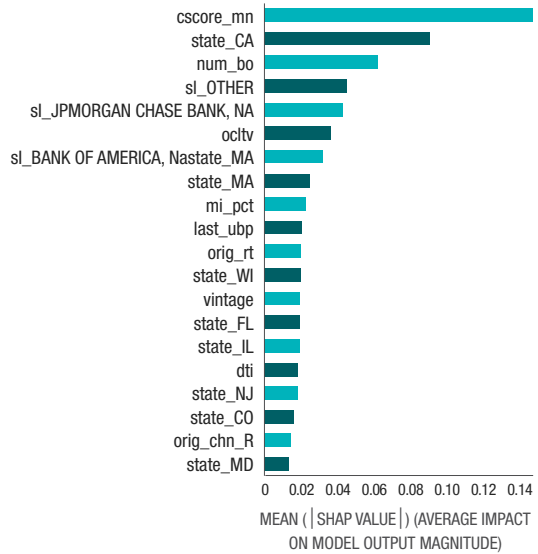
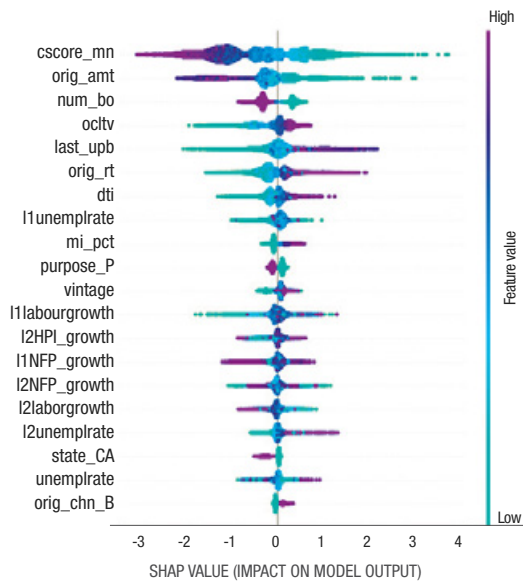


Figure 2: Shapley values based on XGBOOST



Data scientists have frequently come from quite heterogeneous backgrounds and with significant differences from econometricians. Using data from LinkedIn, Stitch Data (2015) summarizes the background of data scientists and finds that computer science is the most common background.

Segaran and Hammerbacher (2009) has an interesting article by Jeff Hammerbacher, who supposedly coined the term “data scientist” while leading the data team at Facebook on the eclecticism in the data science backgrounds.

In this article, we focus on some elements of inference around analytics possible in machine learning, contrasting them with how applied econometricians traditionally approached inference. We do this in the context of applying both traditional econometric methods and several machine learning methods to the same data set.

This is the publicly available FNMA 30-year fixed rate mortgages. We then compare and contrast what drivers of risk are identified using some traditional econometric methods as well as different machine learning methods.

Parallels to statistical inference in machine learning/deep learning models are currently focused very heavily on the twin concepts of interpretability/explanability. Most commonly promoted explainability metrics have been Shapley value and feature importance. For example, the AI platforms of both Google and Microsoft provide Shapley values for users to understand what drives the models as well as to identify model bias. Vendors, such as ZestFinance and DataRobot, have also promoted Shapley value as the way to “break open

³ Merrill, D., 2019, “CEO ZestFinance, Testimony to the House Committee on Financial Services AI Task Force,” June 26, <http://tinyurl.com/y845wptd>

the blackbox.”³ The academic literature on machine learning has also focused on significance tests based on Shapley values and/or feature importances. For example, Horel and Giesecke (2020) develop an asymptotic theory for neural networks using gradients from the fitting algorithms.

Our exploratory analysis in this paper shows that different state-of-the-art machine learning methods can produce models that are similar in their predictive abilities. However, commonly used interpretability metrics can lead to different conclusions about the key risk drivers.

2. DATA

We use the Single-Family Historical Loan Performance Dataset from FNMA. We select the loans originated in the years 2000, 2001, and 2002. The outcome we model is the probability of a loan becoming 90 days past due in the five years after origination. We also combine the national level macroeconomic variables HPI Index, Unemployment Rate, Labor Force, and Non-farm Payroll. These are expressed as growth rates and their first two lags are used. For the categorical variables, we create dummies, or what is referred to in machine learning as one-hot encoding.

3. RESULTS

We first apply two most commonly used machine learning algorithms, XGBOOST and lightGBM, and secondly deep learning with Keras. We optimize the hyper-parameters by grid search. The performance of the three algorithms, as measured by the area under the curve (AUC), is quite similar. We then plot the Shapley values for the features in three figures. These are for lightGBM in Figure 1, XGBOOST in Figure 2, and Keras in Figure 3.

As can be seen in these figures, there are very significant overlaps between the three methods. However, there are also important differences. The two methods, lightGBM and XGBOOST, broadly select the same set of borrower characteristics in the top five. However, XGBOOST selects lagged unemployment rate as the eighth most significant driver. In contrast, lightGBM does not have any macroeconomic variables in the top ten drivers. Deep learning via Keras has a very different set of features selected as the most important ones based on Shapley values.

We then use econometric methods to identify what drives default. We choose the Elasticnet method, which was proposed by Zou and Hastie (2005) and has been used in more than 20,000 studies. The Elasticnet method bridges the “least absolute shrinkage and selection operator” LASSO method and ridge regression.

$$\min \|y - X\beta\|^2 \text{ subject to } \sum_{j=1}^m |\beta_j| \leq t_1, \sum_{j=1}^m \beta_j^2 < t_2$$

Elasticnet ends up with 16 variables or features. We then run a logistic regression with the selected features. The results are presented in Table 1. These results show that Elasticnet finds significantly greater importance for the macroeconomic variables. Lagged Nonfarm Payroll growth and unemployment rate show up as the third and fourth most important variables.

4. CONCLUSION

Using single family mortgage data, we find that different machine learning algorithms can produce rather different rankings of the variables that drive the outcome of interest. This suggests that one needs to exercise caution in relying on these methods in terms of identifying the drivers of risk.

Table 1: Results from estimating logistic regressions for mortgage delinquency

VARIABLE	PARAMETER	STD. ERROR.	WALD χ^2
Intercept	4.860	0.176	759.02
cscore_mn	-0.016	0.000	16070.73
l1NF growth	5.541	0.600	85.42
l2unemprate	0.028	0.005	36.02
mi pct	0.013	0.001	251.29
numbo	-0.791	0.015	2740.16
ocltv	0.018	0.001	433.90
orig amt	0.000	0.000	975.22
orig chn B	0.208	0.019	117.21
orig chn R	-0.066	0.018	13.69
orig rt	0.383	0.015	634.05
prop typ CO	-0.376	0.046	66.71
prop typ CP	-0.605	0.128	22.46
prop typ MH	0.813	0.057	201.27
prop typ SF	0.101	0.032	10.06
purpose P	-0.591	0.022	700.70
purpose R	-0.028	0.024	1.35

This table reports the parameter estimates from a logistic regression of key drivers of mortgage delinquency that had been identified via an Elasticnet regression. The sample had been divided into 80% training and 20% validation subsamples. The variables are first selected via an Elasticnet method. A logistic regression is run with the top 21 selected variables and the results are presented below. The out of sample AUC is 0.852.

REFERENCES

- Biddle, J., 2017, "2016 Hes Presidential Address: Statistical inference in economics, 1920–1965: changes in meaning and practice," *Journal of the History of Economic Thought* 39, 149–173
- Butaru, F., Q. Chen, B. Clark, S. Das, A. W. Lo, and A. Siddique, 2016, "Risk and risk management in the credit card industry," *Journal of Banking & Finance* 72, 218–239
- Horel, E., and K. Giesecke, 2020, "Significance tests for neural networks," *Journal of Machine Learning Research* 21, 9291–9319
- Sadhvani, A., K. Giesecke, and J. Sirignano, 2021, "Deep learning for mortgage risk," *Journal of Financial Econometrics* 19, 313–368
- Segaran, T., and J. Hammerbacher, 2009, "Beautiful data: the stories behind elegant data solutions," O'Reilly & Associates Inc.
- Stitch Data, I., 2015, "The state of data science," <http://tinyurl.com/eurxb74w>
- Zou, H., and T. Hastie, 2005, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67, 301–320

APPENDIX

Table A1: This table lists the features (i.e., variables) from the FNMA data

FEATURE	DESCRIPTION
cscore_mn	Borrower credit score; FICO score.
last_upb	The current actual outstanding unpaid principal balance of a mortgage loan, reflective of payments actually received from the related borrower.
mi_pct	The original percentage of mortgage insurance coverage obtained for an insured conventional mortgage loan and used following the occurrence of an event of default to calculate the insurance benefit.
mi_type	"The entity that is responsible for the Mortgage Insurance premium payment. 1 = borrower paid; 2 = lender paid; 3 = enterprise paid; * Null = No MI"
num_bo	The number of individuals obligated to repay the mortgage loan.
num_unit	The number of units comprising the related mortgaged property (one, two, three, or four).
occ_stat	The classification describing the property occupancy status at the time the loan was originated. Principal = P; second = S; investor = I; unknown = U
ocltv	The ratio, expressed as a percentage, obtained by dividing the amount of all known outstanding loans at origination by the value of the property.
orig_amt	Origination amount
orig_chn	Origination channel: retail = R; correspondent = C; broker = B
orig_rt	The original interest rate on a mortgage loan as identified in the original mortgage note.
orig_trm	Original term
prop_typ	"Property type: CO = condominium CP = co-operative PU = Planned Urban Development MH = manufactured home SF = single-family home"
purpose	"An indicator that denotes whether the mortgage loan is either a refinance mortgage or a purchase money mortgage. Cash-Out Refinance = C Refinance = R Purchase = P Refinance-Not Specified = U"
dti	The ratio obtained by dividing the total monthly debt expense by the total monthly income of the borrower at the time the loan was originated.

IMPLEMENTING DATA GOVERNANCE: INSIGHTS AND STRATEGIES FROM THE HIGHER EDUCATION SECTOR

PATRICK CERNEA | Director, Data Strategy and Governance, York University, Canada

MARGARET KIERYLO | Assistant Vice-President, Institutional Planning and Chief Data Officer, York University, Canada

ABSTRACT

This article explores the critical role of data governance in the context of higher education. The authors highlight the strategic importance of establishing comprehensive data governance frameworks to enable data-informed decision-making and argue that data governance can enhance strategic enrollment management, efficiency and effectiveness, and enable innovation. The authors present a detailed exploration of the strategies for building data governance capabilities within higher education institutions. They outline the process of setting data governance goals, selecting operating models for data governance, considering resourcing models, defining roles and responsibilities, establishing data governance committees, and identifying metrics to assess progress. Practical applications of data governance, including metadata management, data quality management, and ensuring regulatory compliance and ethical use of data, are discussed to illustrate how institutions can enhance their data environment. The authors conclude by exploring future trends and emerging issues in data governance within higher education, pointing to the lag in data governance advancement compared to other sectors and the imperative for post-secondary institutions to adopt robust data governance frameworks to remain competitive and innovative.

1. INTRODUCTION

1.1 The data environment

Organizations – including institutions of higher education – agree that data is critical to management and decision making. Over the past several decades, digitalization and the use of data have expanded into every corner of higher education institutions, including universities and colleges. Departments have become increasingly data-driven, relying on data for their daily operations. As the number of data creators and users has grown, so too has the variety of approaches to collecting, producing, distributing, and analyzing data. A fully data-driven post-secondary institution extends beyond using data for routine tasks, embedding data-derived insights into strategic and operational decision making at all levels, from front-line units to senior leadership.

Despite the widespread acknowledgment of the importance of these developments, many institutions are still at the formative stages of developing and implementing comprehensive data governance frameworks. In Canada, for example, a 2023 survey conducted by the Canadian Association of University Business Officers (CAUBO) revealed that 47% of higher education institutions are just beginning their data governance journeys [Al-Hussein et al. (2023)]. This statistic highlights that nearly half of Canadian post-secondary institutions are beginning to realize the importance of systematically managing their data assets.

Using data well is a core competency for all successful organizations, and all sectors must continually get better at it. In the post-secondary context, beyond enabling reporting, institutions are leveraging data to grow enrollment, enhance efficiency and effectiveness, and drive innovation. This

strategic use of data for growth and efficiency sets the stage for adopting advanced technologies, notably artificial intelligence (AI). AI algorithms and machine learning techniques are now being employed to analyze complex datasets, offering insights that previously required extensive effort. These technologies are enabling post-secondary institutions to predict trends, automate tasks, and personalize student experiences. This evolution in data utilization marks a significant shift in how institutions approach their strategic goals, illustrating the dynamic nature of data as a tool for impact and innovation.

Data is fundamental in how we prepare for the future, share knowledge, and make good decisions. However, transitioning to a fully data-driven model entails significant planning and coordination. For many organizations, the full potential of data and analytics capabilities remains untapped. The key requirement is not simply more data or more assessment, but the establishment of systematic data and analytics practices.

In this context, the formulation of comprehensive data strategies is vital for organizations to harness the power of their information assets effectively [Powers (2020)]. These strategies should not only outline the overarching framework for data utilization but also emphasize the development of key institutional functions: data literacy and data governance. By enhancing data literacy, organizations equip their community with the skills necessary to interpret, analyze, and leverage data effectively. Simultaneously, data governance ensures the integrity, security, and ethical use of data [Eryurek et al. (2021)]. Together, these elements create a strong foundation for a data-driven culture, one that facilitates informed decision making and fosters innovation across all levels of the organization. In the post-secondary context, the aim is to transform data into a strategic asset that enhances teaching and learning, improves student experiences, allows for the delivery of smarter student

services, enables effective strategic enrollment management, promotes operational effectiveness, excellence, and efficiency, and drives institutional innovation.

1.2 What is data governance?

Data governance encompasses the practices, policies, processes, standards, and metrics required to manage data as an asset [DAMA (2017)]. Its purpose is to utilize data effectively and efficiently in support of organizational goals. Data governance has a symbiotic relationship with data management: data governance establishes data policies and procedures, while data management implements those policies and procedures to aid in decision making.

We can illustrate this distinction by using data quality as an example:

- Data governance involves establishing data quality standards (e.g., the acceptable levels of data accuracy, completeness, and consistency), specifying the roles and responsibilities related to data quality (e.g., the responsibilities of “data stewards” and “data custodians”), and creating policies and processes that dictate how data quality should be maintained and enhanced.
- Data management involves the practical implementation of these policies, standards, and procedures. This includes day-to-day operations and technologies used to ensure data meets established quality standards. Activities under data management may include data quality controls (e.g., using software tools to monitor and report on data quality), conducting data quality audits, and undertaking data cleansing and improvement efforts.

An overview of data governance and its relationship to a typical university planning framework is provided in Figure 1.

Figure 1: Data governance in a typical university planning framework



1.3 Key data governance terms

To provide a foundational understanding of data governance, it is important to understand the core terms and concepts that make up this field. This section delves into key data governance terminologies, offering insights into their meanings and implications within an organization. Key data governance terms include:

- **Data:** refers to all quantitative and qualitative information that is collected, stored, managed, analyzed, and utilized across various organizational departments and functions.
- **Data custodian:** generally, the individual who is responsible for the technical management and security of a particular data system or dataset.
- **Data domain:** generally, a specific category or subject area of data within the organization. It is a broad area of data that contains a set of similar or related data elements, such as financial data and human resources data.
- **Data steward:** generally, a senior manager who is responsible for the data in one or more data sub-domains. Data stewards are usually required to be experts in data within their data sub-domain(s).
- **Data sub-domain:** generally, a subset or a specific aspect of a data domain. It is a smaller, more specific area of data (that is part of a larger data domain), such as staff profile data, payroll data, and employee safety data.
- **Data trustee:** generally, a senior executive who is accountable for the data in one or more data domains. Data trustees usually have decision-making authority regarding the authoritative sources of data that are managed and created by the central unit.
- **Metadata:** structured, descriptive information about data elements and data assets that provides context, facilitates understanding, and enables effective management, discovery, and usage. For instance, for a "student ID" data element, the metadata might include a definition and a validation rule.
- **Principal data:** core data that is essential for the organization's operations and decision-making processes. This data is used to identify, describe, and manage the primary aspects of the organization. At higher education institutions, principal data includes information about students, alumni, staff, faculty, academic programs and services, organizational and financial structures, and physical space. Principal data is often synonymous with "master data".

- **Reference data:** the sets of predefined, permissible values or categories that are used within the organization's systems and databases to classify, organize, and ensure the consistency of data. This data provides context and structure to transactional and operational data, enabling accurate data interpretation, reporting, and analysis. At higher education institutions, reference data includes country codes, currency codes, and program classification codes.

1.4 Making the case for data governance

Data governance facilitates better decision making and operational efficiency and effectiveness primarily through improved metadata, data quality, data protection and compliance, and other data management policies and procedures. There are two main reasons to embrace data governance: increasing the value of data and reducing risks associated with poor data management.

1.4.1 INCREASING THE VALUE OF DATA

Today's data environment, characterized by growing volume, complexity, and AI breakthroughs, makes data governance essential for organizations to maintain or gain a competitive advantage. Enhanced visibility over data assets, increased data literacy, standardized data language, and improved data quality all contribute to making data more valuable, thus improving business outcomes. Companies such as Airbnb, GE Aviation, and Uber all leverage data governance to enhance decision making [Atlas (2023)]. In a 2023 survey of the Canadian higher education sector, 41% of universities and colleges highlighted supporting decision making as a key outcome of data governance at their institution [Al-Hussein et al. (2023)].

1.4.2 REDUCING RISKS

The evolving data landscape brings increased regulatory requirements, exemplified by the European Union's General Data Protection Regulation (GDPR), which highlights the legal implications of data mismanagement and risks like security breaches, revenue loss, and reputational damage [E.U. (2016)]. Data governance is crucial for risk management. In the Canadian higher education sector, 18% of universities and colleges highlighted ensuring compliance as a key outcome of data governance at their institution [Al-Hussein et al. (2023)].

2. BUILDING DATA GOVERNANCE CAPABILITIES

This section highlights key considerations in the development of data governance programs. Outlined below are five principles that guided the implementation of data governance at our university.

- **Strategy should drive data governance efforts:** data strategies should inform data governance processes as they provide a comprehensive roadmap to effectively manage and leverage data assets. By aligning data governance processes with an overall data strategy, institutions can ensure that their data governance efforts are focused and relevant to their specific needs and objectives. In the Canadian higher education sector, 21% of universities and colleges highlighted the lack of a business case as a roadblock to data governance adoption [Al-Hussein et al. (2023)].
- **Communication is essential:** since data governance fundamentally involves people, institutions often find that it is crucial to focus on clear and frequent communication. Effective communication strategies are vital to successfully implement data governance, as they ensure that stakeholders are aligned and engaged with institutional data governance initiatives.
- **Data governance efforts should be focused:** institutions should start their data governance journey by focusing on key data domains and initiatives. Data governance is often misunderstood or seen as a high-level strategic initiative, making quick wins essential. Focusing data governance efforts also streamlines resource allocation. In the Canadian higher education sector, 45% of universities and colleges highlighted capacity risk as a roadblock to data governance adoption [Al-Hussein et al. (2023)].
- **Data literacy should be prioritized:** data literacy is crucial for effective data governance. All members of the organization, regardless of their role or unit, should develop a basic understanding of data and its significance. This facilitates better collaboration, informed decision making, and more effective data governance practices. In the Canadian higher education sector, 55% of universities and colleges highlighted a lack of data literacy as a roadblock to data governance adoption [Al-Hussein et al. (2023)].

- **Data governance implementation requires robust change management:** institutions should pay special attention to change management, especially in the early stages of data governance implementation when goals and deliverables are ambiguous. In the Canadian higher education sector, 58% of universities and colleges highlighted a lack of change management as a roadblock to data governance adoption [Al-Hussein et al. (2023)].

2.1 Data governance vision, mission, and goals

Organizations typically begin their data governance journey by crafting a vision and mission for the program, ensuring alignment with their overarching data strategy. Although each organization is unique, Canadian higher education institutions often have vision and mission statements that emphasize leveraging data as a strategic asset, ensuring data quality and security, and fostering data-informed cultures within their institutions.

The data governance program derives its goals from its vision and mission. In the context of Canadian higher education institutions, these goals commonly prioritize enhancing data quality, maintaining metadata, developing data literacy, ensuring compliance, fostering a culture of data sharing, and protecting data assets. These goals help operationalize data governance initiatives and facilitate tracking progress against targets.

2.2 Operating models for data governance

Implementing an operating model that aligns with the organization's unique data environment and business objectives is paramount. Centralized, decentralized, and federated models each have distinct advantages and challenges.

- **Centralized data governance:** typically consolidates data governance authority in one department or unit. Although this model may benefit from uniformity and consistency, it may suffer from slower decision making and a lack of functional subject-matter expertise.
- **Decentralized data governance:** distributes authority across the organization and various departments. This model promotes agility and customization but may lead to inconsistencies around data handling, data definitions, and the implementation of policies and processes.

- **Federated data governance:** blends centralized oversight with decentralized execution. It offers a compromise between uniformity and flexibility. This model works well for most organizations, including higher education institutions.

At our university, we implemented a federated data governance model in which the central data governance team, in consultation with key partners such as data trustees and data stewards, defines policies, standards, and guidelines for data management. The central data governance team is also responsible for managing the institutional metadata repository and developing the university's data literacy program. Day-to-day data management activities are handled at the sub-domain level by data stewards and data custodians, with support from information technology (IT).

2.3 Resourcing models for data governance

Resource allocation varies by organization size and needs. Organizations typically consider in-house teams, outsourcing, or hybrid models based on available resources and expertise. Decisions should be guided by the nature and duration of the data governance work. If an organization prioritizes data quality, does it have the required expertise to undertake that work? If not, can it hire that expertise? The duration of the work also plays a crucial role: short-term projects might be better suited for outsourcing, whereas long-term engagements may benefit from developing and retaining expertise internally. For example, organizations might choose to do the metadata work in-house and outsource for data quality assessments and cleansing.

2.4 Roles and responsibilities in data governance

With the resourcing model in place, organizations must consider additional roles and responsibilities related to data governance. Stakeholder maps and engagement plans allow for the identification and involvement of key groups in data governance activities.

As data governance is often perceived as yielding less tangible results, organizations typically begin their journey by securing an executive sponsor to advocate for the data governance program. Another crucial role is the chief data officer (CDO), who is responsible for overseeing the data governance strategy, ensuring data quality, and driving the cultural change toward data-driven decision making across the organization.

Organizations subsequently assign accountability for enterprise data by segmenting it into data domains and sub-domains. Typically, in the higher education sector, data domains represent broad categories like student, financial, and human resources data, while sub-domains are more specific areas within these domains.

Generally, as noted in Section 1.3 above, there are three roles found in higher education institutions:

- **Data trustees:** usually senior executives who oversee one or more data domains.
- **Data stewards:** often senior managers who manage one or more data sub-domains and are considered experts in their areas.
- **Data custodians:** typically IT professionals who handle the technical aspects of data sub-domains, systems, or both.

Fundamentally, data governance hinges on people and cross-functional collaboration for success and it permeates the entire organization.

2.5 Data governance committees

Data governance committees are essential, as they facilitate decision making and policy development and help advance operational work. The specific composition of data governance roles may require creating dual committees alongside various working groups:

- A highly strategic data trusteeship committee or data governance council that sets the direction for the data governance program and ensures alignment with the institutional data strategy.
- A more operational data stewardship committee that plans and undertakes project work.
- Working groups that reflect data sub-domains, involve data custodians, or both.

Whatever structure is established, committees should have clear mandates and goals, involve all relevant stakeholders, and ensure consistency through regular meetings.

2.6 Metrics: Measuring the effectiveness of data governance

Data governance is often underprioritized as an institutional function because of two key issues: misalignment with strategic and operational objectives and a lack of tracking data governance initiatives. Identifying key performance indicators for each data governance goal is crucial for monitoring progress and assessing the impact on business operations.

Metrics used to track the progress of data governance goals might include the number of data stewards identified, the number of data definitions approved, and measures of engagement in data governance workshops.

Metrics that measure the impact on operations are harder to quantify, but they are ultimately critical to showcase value. They can include efficiency gains due to data governance, decreased penalties from avoiding regulatory non-compliance, or time saved due to improved data quality [for examples of other metrics, see Pansara (2023)].

Data governance measures should be actionable and top-of-mind for those accountable for data domains and sub-domains, and progress should be shared with institutional stakeholders.

3. THREE PRACTICAL APPLICATIONS

With foundational data governance elements established, organizations can start to undertake data governance work to enhance their data environment. Three practical applications include metadata management, data quality management, and regulatory compliance and ethical use of data. It is important to note that implementing these initiatives requires collaboration across various teams. Specifically, improving data quality requires significant collaboration between the data governance and data management teams and business units.

To prioritize data governance projects and sub-domains of focus, organizations typically conduct a business impact analysis to understand the value and sensitivity of their data assets while identifying high-risk areas. An instrumental part of this process is the development of a prioritization matrix that incorporates criteria such as strategic and operational alignment, business value, revenue potential, risk mitigation, and resource availability.

Higher education institutions typically start this prioritization process by identifying the data most critical to key institutional initiatives, such as strategic enrolment management. In the Canadian higher education sector, 68% of universities and colleges highlighted performance data collection and analysis, and 63% highlighted enrollment management as key use cases of data governance at their respective institutions [Al-Hussein et al. (2023)]. Often, this involves prioritizing student sub-domains such as “student profile”, “recruitment and admissions”, and “student advising”, as well as principal and reference sub-domains such as “principal academic programs and services” and “reference geographic locations”. The work undertaken in those sub-domains often begins with a focus on managing metadata, ensuring data quality, promoting data literacy, and developing relevant data policies and procedures.

3.1 Metadata management

As noted in Section 1.3, “metadata” is information that describes and provides context for other data. In essence, it describes the various aspects of data, like its content, format, source, and context. Organizations gain visibility and understanding of their data once they inventory it, define it, and track its lineage. Metadata management, therefore, involves ensuring that data is managed with the same rigor as any other valuable asset. A 2023 survey of post-secondary institutions found that only 25% of respondents reported that their institutions have clear and comprehensive data definitions that adequately cover the nuances of data [Al-Hussein et al. (2023)].¹

Why should organizations undertake metadata management? Consider the following scenarios:

- Senior management does not fully grasp the intricacies of the data in an institutional performance report. The solution might be to define data elements in an institutional metadata repository.
- An organization is implementing a new customer relationship management (CRM) software and notices unnecessary duplication of data assets. The first step might be to catalog data assets, their content, formats, and sources.
- An organization wants to improve the quality of its financial data but does not know where to start. The first step might be to inventory data elements, enabling them to document data quality rules and conduct a data quality assessment.

¹ Results include “strongly agree” and “somewhat agree” out of a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

Given that metadata management is crucial for the proper governance of an organization's data assets, applying a set of best practices will facilitate its successful implementation. The following are recommendations developed through operationalizing metadata management at our university:

- **Establish clear policies and standards:** a metadata guidelines document, for instance, should define the types of metadata accepted, prescribe standards for writing definitions, and set standards for cataloging data assets.
- **Involve stakeholders across departments and units:** metadata management should involve data trustees, data stewards, data custodians, and subject matter experts. In creating institutional data definitions, it is essential to include relevant experts. For instance, when standardizing reference geographic data, institutions should involve those who manage and use this data to ensure a consistent organizational standard.
- **Provide training opportunities and conduct awareness campaigns:** educate staff on the importance of metadata management and its effective uses. Institutions could decide to provide every data trustee, data steward, data custodian, and subject matter expert with an onboarding session on the institutional metadata repository.
- **Use metadata management tools:** specialized tools and software such as Informatica Enterprise Data Catalog, Collibra, and Data Cookbook are specifically designed to help create, store, and retrieve metadata effectively. Tools should be chosen based on the size and complexity of the institution and its budget.
- **Enforce data security and privacy:** institutions should ensure their metadata management practices comply with security and privacy policies. Additionally, they should use their metadata guidelines document and onboarding sessions to communicate these standards. For instance, data definitions should not include sensitive information.

We began our metadata management journey by defining terms required for strategic enrollment management and our digital transformation program. This meant primarily involving teams from the “student” and “principal academic programs and services” domains and ensuring integration of data definitions with relevant dashboards and reports, such as the university's enrollment management dashboard.

3.2 Data quality management

High-quality data leads to better decisions, facilitates strategic planning, and reduces the time employees spend on ad-hoc assessments, data manipulation, and cleansing. Implementing transparent processes for managing data quality can also significantly enhance trust in organizational data. A recent survey of Canadian post-secondary institutions found that only 66% of respondents believe their institutions' data is trustworthy [Al-Hussein et al. (2023)].

The key to improving data quality sustainably is to establish an institutional data quality management program. This approach allows organizations to focus their scope, align data quality improvements to business outcomes, and streamline resource allocation. In essence, this enables organizations to:

- Enhance decision making and drive strategic impact by improving the value and usability of their data.
- Increase efficiency and productivity by streamlining data processes and minimizing delays caused by data inaccuracies.
- Improve customer and stakeholder satisfaction through reliable data-driven services and interactions.
- Reduce risks associated with poor data quality, such as compliance issues and reputational damage.
- Enable a modern ecosystem of integrated information platforms and applications, which requires high-quality data to function properly.

Data quality improvements should have a specific and focused scope. Since data quality efforts can become costly and time-consuming, organizations should initially aim for quick wins and impact on critical business areas. Typically, this is done by surveying data producers and users to identify the most significant data quality issues. Improvements should then be prioritized based on business value. Higher education institutions usually begin their data quality journey by addressing student data. According to the Data Management Body of Knowledge, a systematic approach to data quality includes:

- Defining high-quality data.
- Defining a data quality strategy.
- Identifying critical data and business rules.
- Performing an initial data quality assessment.
- Identifying and prioritizing potential improvements.

² Ibid.

- Defining goals for data quality improvement.
- Developing and deploying data quality operations including:
 - managing data quality rules
 - measuring and monitoring data quality
 - developing operational procedures for managing data issues
 - establishing data quality service level agreements
 - developing data quality reporting [DAMA (2017)].

Building on this systematic approach, enhancing an organization's data quality typically involves implementing a set of best practices tailored to its specific needs and goals. To ensure the effectiveness of their data quality management program, organizations should:

- Prioritize data quality work based on sub-domains and systems critical to the business.
- Leverage use cases and data quality process maps as catalysts for data quality.
- Ensure adequate resources are in place in IT to enable data management functions.
- Implement measures to keep low-quality data out of the organization's data ecosystem. This often involves establishing data entry controls and defining data quality rules.
- Leverage tools for data quality profiling (e.g., IBM InfoSphere Information Analyzer, Collibra), modeling and “extract, transform, and load” (ETL) processes (e.g., Informatica, AWS Glue), metadata management (e.g., Informatica Enterprise Data Catalog, Data Cookbook), incident tracking (e.g., Jira, Zendesk), and data quality reporting (e.g., Power BI, Tableau).
- Empower stakeholders responsible for data quality (e.g., data stewards) to:
 - decide if their data is sufficiently complete and accurate to support business process needs
 - set up targets for specific attributes
 - set up thresholds for the level of quality acceptable
 - establish measures and metrics to track improvements [HealthIT].

Our university began enhancing data quality in the student information system by prioritizing the data assets and elements critical to strategic enrollment management. This effort primarily involved data stewards and data custodians from the “student profile” data sub-domain. Initial tasks included

documenting existing data quality processes, capturing data quality rules in the institutional metadata repository, and performing an initial data quality assessment.

3.3 Regulatory compliance and ethical use of data

In the context of data management, regulatory compliance and ethical considerations in data usage are of paramount importance, especially with the rise of technologies like AI and cloud computing, which have increased the possibilities for innovative (and unethical) data use. Adhering to legal standards and ethical guidelines is essential not only to avoid legal repercussions but also to maintain public trust and safeguard the rights and privacy of individuals. For readers of this article, various legislation and regulations could apply, including:

- the General Data Protection Regulation (GDPR) in the European Union
- the California Consumer Privacy Act (CCPA) in the United States
- international standards like ISO/IEC 27001 for information security management
- sector-specific regulations, such as those in finance and education (e.g., Ontario's Ministry of Training, Colleges and Universities Act).

Ethical use of data goes beyond legal compliance; it encompasses respect for confidentiality, consent, fairness, and transparency in data handling. A commitment to ethical data usage is crucial in building a responsible and sustainable data culture within organizations.

3.3.1 ETHICAL PRINCIPLES FOR THE USE OF DATA

At our university, this work was undertaken by the Principles for the Ethical Use of Student Data Working Group, which had representation from across the university. The group was tasked with developing guiding principles to ensure ethical use of student data, aligning with the university's commitment to decolonization, equity, diversity, and inclusion (DEDI), and compliance with institutional values and policies.

Between Fall 2022 and Spring 2023, the working group focused on developing high-level principles for future projects and activities involving student data. They balanced various values, such as student privacy and the duty to act, and recognized the ongoing debate around ethical principles. The group emphasized the need for a continuous, fact-informed discussion.

The scope of the principles covers all data related to current and prospective students, and alumni. It includes a wide range of activities from advising to the use of AI and learning analytics.

Key principles include:

- **Consent:** this principle refers to the explicit, informed, and meaningful agreement given by an individual for their personal data to be collected, processed, or analyzed for a specific purpose.
- **Transparency:** this principle refers to the obligation to be open and honest about organizational data practices. The principle of transparency facilitates the ability of students and alumni to provide free and informed consent.
- **Duty of care:** this principle refers to the obligation to take steps to ensure that data is collected, processed, and used in a way that does not cause harm.
- **Obligation to act:** this principle is about ensuring that personal student data is collected, processed, and used in a way that is aligned to the best interests of students.
- **Data minimization:** this principle refers to collecting, processing, and using the minimum amount of personal data necessary to achieve a specific purpose.
- **Stewardship of data:** this principle refers to the responsible management of data.

These principles reflect emerging practices for ensuring compliance and ethical data usage. To ensure continued alignment with best practices, the working group's recommendations include developing resources for faculty and staff to build literacy in interpreting student data, conducting an annual review of the principles, and implementing a communication plan. The operationalization of the principles is currently underway. For instance, when making data requests, the requestor is required to review and align their request with the "Principles for the ethical use of student data". This step ensures compliance with the established ethical guidelines set forth for student data usage.

3.3.2 METHODS FOR PRIORITIZING COMPLIANCE AND ETHICAL USE OF DATA IN BUSINESS OPERATIONS

As AI continues to permeate various sectors including finance, transportation, healthcare, and higher education, it will become increasingly important to balance these technologies against potential ethical risks [Kaushikkumar (2024)]. To effectively prioritize ethical data usage within an organization's operations, organizations need to adopt a strategic approach.

Formulating comprehensive policies, guidelines, or frameworks is essential; these should meet legal standards and embody ethical principles that align with the organization's strategic directions while balancing benefits and risks.

Furthermore, it is essential to initiate regular training programs for employees, focusing on data ethics and legal compliance. These training sessions are crucial in cultivating a culture of awareness and responsible data usage.

Another key method is conducting thorough data audits. These audits play a vital role in verifying adherence to both regulatory requirements and ethical principles, thereby reinforcing overall compliance.

Finally, the integration of ethical considerations into decision-making processes, especially for projects involving personal data, is critical. This practice ensures that decisions are made with an ethical lens, not just a legal one, thus embedding a sense of trust and integrity in business operations.

Together, these methods form a strong framework for ensuring that compliance and ethical use of data are central to an organization's operations [Braunack-Mayer et al. (2020)]. The integration of ethical considerations into data usage is not just a legal necessity; it is a cornerstone of building trust and integrity in business operations. By prioritizing these aspects, organizations not only protect themselves from legal risks but also establish themselves as responsible and trustworthy entities in the eyes of their customers and the public.

4. CONCLUSION AND FUTURE TRENDS

Data governance represents a fundamental shift in how organizations value and manage one of their most critical assets: data. This article has provided insights into data governance within the context of higher education. As we have demonstrated, data governance is a critical function as institutions prepare for the future of learning; insights derived from well-governed data can lead to transformative outcomes for students, faculty, and the broader community. Without data governance, an institution's ability to leverage data as a strategic asset is limited. As we have explored, the data governance journey is multifaceted and dependent on collaboration and coordination, and involves comprehensive metadata management, a rigorous pursuit of data quality, and adherence to ethical standards.

Through the establishment of comprehensive data strategies, higher education institutions can harness the full potential of their data, enhancing decision making, operational efficiency

and effectiveness, and innovation. The journey involves not only the technical aspects of data management but also a cultural shift towards data literacy and a shared responsibility for data governance across all levels of the institution.

As we look to the future, several trends and emerging issues are evident:

- Compared to sectors such as health and finance, post-secondary institutions have lagged in advancing data governance as a core competency. Universities and colleges will seek to unlock the full potential of their data assets through the implementation of data governance frameworks [CAUBO (2023)].
- The pace of technological innovation may outstrip the ability for regulatory frameworks to adapt. AI and machine learning rely heavily on large datasets. Data governance ensures that the data used to train these systems is not only high quality and relevant but also ethically used. This underscores the importance of data governance as organizations navigate the complexities of an AI-driven future [Kaushikkumar (2024)]. As regulations evolve to catch up with technology, organizations with robust data governance frameworks will be better positioned to remain agile and competitive in their respective fields.
- Given identified research gaps in data governance, future research trends will explore data privacy and the evolving landscape of data governance, with emphasis on the interplay between AI and data stewardship practices. Other topics of interest include organizational challenges related to governance implementation, the impact of AI on data governance, and cross-border regulatory compliance [Pansara (2023)].

- Concerns about bias in “AI algorithmic decision support” will continue. Data governance can assist by ensuring that datasets are diverse and representative. By monitoring and managing the composition of datasets (metadata), data governance can help anticipate and prevent biases that arise from underrepresented groups or skewed data samples, leading to more equitable AI outcomes [Davidson (2023)]. Ethical considerations will become as important as legal compliance.
- Institutions committed to decolonization, equity, diversity, and inclusion (DEDI) and Indigenization face the challenge of navigating an even more complex data governance landscape. Understanding Indigenous ways of knowing and Indigenous data systems will be crucial in supporting Indigenous data sovereignty [Animikii, 2022].

As we have demonstrated, the role of data governance in higher education will only continue to grow in importance. Implementation requires a commitment to continuous improvement, collaboration, and alignment with institutional priorities. Moreover, data governance efforts should be focused, targeting areas of greatest impact and importance. Prioritizing data literacy is essential, as it empowers individuals across the institution to effectively interpret data and apply insights to enable informed decision making. Finally, change management is required to navigate the complexities of data governance implementation successfully. By prioritizing data governance and committing to effective data management, post-secondary institutions can ensure they are well-positioned to meet the challenges of the 21st century, driving innovation and excellence in higher education.

REFERENCES

- Al-Hussein, R., P. Cernea, Z. Chan, M. Kierylo, and M. Morgado. 2023, “Data governance capabilities,” Webinar, Canadian Association of University Business Officers, October 12, <http://tinyurl.com/2y2dkrf3>
- Animikii, 2022, “#DataBack: asserting and supporting indigenous data sovereignty,” <http://tinyurl.com/yfnu7h7r>
- Atlan, 2023, “5 data governance examples: case studies, takeaways and more.” May 25, <http://tinyurl.com/2v5uv6xn>
- Braunack-Mayer, A. J., J. M. Street, R. Tooher, X. Feng, and K. Scharling-Gamba, 2020, “Student and staff perspectives on the use of big data in the tertiary education sector: a scoping review and reflection on the ethical issues,” *Review of Educational Research* 90:6, <http://tinyurl.com/y6b7feps>
- CAUBO, 2023, “New resource: data governance framework for Canadian universities,” Canadian Association of University Business Officers, June 27, <http://tinyurl.com/38anamj3>
- DAMA, 2017, DAMA-DMBOK: Data management body of knowledge, 2nd edition, Technics Publications
- Davidson, E., L. Wessel, J. Sunrise Winter, and S. Winter, 2023, “Future directions for scholarship on data governance, digital innovation, and grand challenges,” *Information and Organization* 33:1, <http://tinyurl.com/67f4dbae>
- Eryurek, E., U. Gilad, V. Lakshmanan, A. Kibunguchy-Grant, and J. Ashdown, 2021, “Data governance: the definitive guide: people, processes, and tools to operationalize data trustworthiness,” O’Reilly Media
- E.U. 2016, “General Data Protection Regulation,” Regulation (EU) 2016/679 of the European Parliament and of the Council, April 27, <http://tinyurl.com/32yzy5j>
- HealthIT, “Data quality assessment,” Office of the National Coordinator for Health Information Technology Health IT Playbook, <http://tinyurl.com/997pb2kh>
- Kaushikkumar, P. 2024, “Ethical reflections on data-centric AI: balancing benefits and risks,” *International Journal of Artificial Intelligence Research and Development (IJAIRD)* 2:1, <http://tinyurl.com/27b2ss83>
- Pansara, R. R., 2023, “Unraveling the complexities of data governance with strategies, challenges, and future directions,” *Transactions on Latest Trends in IoT* 6:6, <http://tinyurl.com/m7sdny4s>
- Powers, K., 2020, *Data strategy in colleges and universities*, Routledge

AI, BUSINESS, AND INTERNATIONAL HUMAN RIGHTS

MARK CHINEN | Professor, Seattle University School of Law

ABSTRACT

This article discusses efforts by policymakers to regulate AI through international human rights. It begins by surveying some of the human rights concerns that arise from AI applications. Because of the important role businesses are playing in the development of AI, the article then sketches the contours of international human rights law as it applies to firms. Businesses have a responsibility to respect human rights, but until recently this has not been understood as a legal obligation. Recent legislation in Europe indicates that the norm is hardening, but there is resistance to this trend. Some of the reasons why are explored here. I join others, however, in arguing that as complex as some of these issues are, international human rights as a set of principles and where appropriate, as legal obligations, are the best overarching framework for governing transformative technologies such as AI.

1. INTRODUCTION

It is now over a year since ChatGPT, the large language model (LLM) developed by OpenAI, galvanized world attention and sparked a race among the major technology firms to deploy LLMs, as well as attempts by policymakers to respond to the risks posed by these applications. The ousting and return of OpenAI's president and the subsequent reorganization of OpenAI's management reflected tensions among the AI community about the directions artificial intelligence (AI) applications should take, the need for capital to develop and monetize them, the influence of large technology companies, and the role of corporate governance in steering AI development and deployment.

It is no surprise that AI applications are subject to such scrutiny, as they have the potential to impact every domain of human life. Public governance of AI is, of course, taking place at the national level, but a nascent form of transnational, regional, and international governance is emerging from the interactions of businesses and private associations, professional organizations, academics, nation states, and international organizations. Such governance comprises a range of soft and hard, technical, and general norms that address AI applications. International human rights are one source of those norms.

This article discusses efforts by policymakers to regulate AI through international human rights. It begins by surveying some of the human rights concerns that arise from AI applications. Next, because of the important role businesses are playing in the development of AI, the article sketches the contours of international human rights law as it applies to firms. Under that law, businesses have a responsibility to respect human rights, but until recently this has not been understood as a legal obligation. Recent legislation in Europe indicates that the norm is hardening, but there is resistance to this trend. Some of the reasons why are explored here. I join others, however, in arguing that as complex as some of these issues are, international human rights as a set of principles and where appropriate, as legal obligations, are the best overarching framework for governing transformative technologies like AI.

2. AI APPLICATIONS AND THEIR IMPLICATIONS FOR HUMAN RIGHTS

AI refers to computer techniques or methods used to perform relatively sophisticated human tasks. It is now being used for diagnostic, predictive, and prescriptive analytics in areas such as transportation, healthcare, the workplace, law enforcement, education, and entertainment. One AI learning technique requires data to train a computer program as it is

being developed for a particular task. Breakthroughs in natural language processing methods, as well as AI models trained on massive amounts of data now allow the generation of images, text, and videos with sometimes startling degrees of realism.

AI applications have the potential to provide significant social and financial benefits. At the same time, observers are concerned that AI applications could lead to adverse impacts in areas such as privacy, safety, democracy, and international peace and security, in turn raising human rights concerns. B-Tech, a United Nations project focusing on human rights and transformative technologies, suggests that nine human rights established under the U.N. Declaration on Human Rights could be negatively affected by generative AI.¹ It raises, for example, the right to privacy set out in article 12 of the Declaration: “No one shall be subjected to arbitrary interference with their privacy, family, home or correspondence, nor to attacks upon their honour and reputation.” B-Tech explains that this right could be violated in several ways. For instance, data used to train generative AI models could contain personal information with no meaningful way for individuals to consent to their collection, particularly if that data is obtained by scraping the web.² Users that interact with AI chatbots could be led to provide personal information without fully understanding how such data will be used.³

AlgorithmWatch is concerned that a lack of transparency around automated decision-making systems “impedes individuals’ access to legal remedies” under article 2(3) of the International Covenant on Civil and Political Rights.⁴ The organization also identifies other risks to human rights such as the right to freedom from discrimination and the rights to freedom of expression, religion, assembly, privacy, and equal treatment.⁵ The E.U. Agency for Fundamental Rights has similarly discussed how several of the 50 rights articulated in the Charter of Fundamental Rights of the European Union could be negatively impacted by artificial intelligence systems.⁶

The international community is now focusing on the human rights implications of business models followed by technology companies. The U.N. Office of the High Commissioner for Human Rights identifies several practices that raise possible concerns:

- Gathering large volumes of personal data (whether to train algorithms or sell insights to third parties);
- Selling products to, or partnering with, governments seeking to use new technologies for state functions or public service delivery that could disproportionately put vulnerable populations at risk;
- The promise of hyper-personalization in human resources or marketing decision[s], which could lead to discrimination;
- Using “algorithmic bosses” to mediate the relationship between workers and firms that generate business value from the offline work being done, while limiting labor protections for those workers;
- Providing a technology that allows vast numbers of small and medium enterprises, or individuals to conduct activities that may result in harm to people, but where control over their activities might be limited; and
- Models that are informed by, or inform, the personal choices and behaviors of populations without their knowledge and consent.⁷

As discussed, the B-Tech project has assessed some of these practices under human rights principles. A detailed factual analysis in a specific case would, of course, be required to determine whether a particular business practice violated a human right as a legal matter, but studies like these confirm that human rights are being used for framing the development and use of artificial intelligence.

¹ U.N., 2023, “Taxonomy of human rights risks connected to generative AI: supplement to B-Tech’s foundational report on the responsible development and deployment of generative AI, Office of the High Commissioner on Human Rights, B-Tech Project, <https://tinyurl.com/4eu7ej89> [hereinafter B-Tech Human Rights Taxonomy]

² Id., at 6. In this regard, see Nasr, M., N. Carlini, J. Hayase, M. Jagielski, A.F. Cooper, D. Ippolito, C. Choquette-Choo, E. Wallace, F. Tramèr, and K. Lee, 2023, “Scalable extraction of training data from (production) language models,” arXiv.org, November 28, <https://tinyurl.com/mr39st9y> (developing a way to attack ChatGPT (gpt-3.5-turbo) so that it disgorges gigabytes of training data, some of which may contain personal information).

³ B-Tech Human Rights Taxonomy, supra note 1

⁴ AlgorithmWatch, 2022, “Position by AlgorithmWatch: Input to the High Commissioner report on the practical application of the United Nations Guiding Principles on Business and Human Rights to the activities of technology companies,” <https://tinyurl.com/4k3fdek7>, page 2

⁵ Id., p. 3, Ashraf, C., 2020, “Artificial intelligence and the rights to assembly and association,” *Journal of Cyber Policy* 5:2, 163-179

⁶ European Agency for Fundamental Rights, 2020, “Getting the future right: artificial intelligence and fundamental rights,” <https://tinyurl.com/2bvzrzuc>; European Agency for Fundamental Rights, 2022, “Bias in algorithms: artificial intelligence and discrimination,” <https://tinyurl.com/4uvhc8tf>. The 50 rights are organized under the headings of dignity, freedoms, equality, solidarity, citizen’s rights, and justice. Charter of Fundamental Rights of the European Union, 2010.

⁷ UN Human Rights Business and Human Rights in Technology Project (B-Tech), 2023, “Applying the UN Guiding Principles on Business and Human Rights to digital technologies: Overview and Scope,” <https://tinyurl.com/mrx7msh>, page 5

3. INTERNATIONAL HUMAN RIGHTS AND BUSINESS

To better understand these trends, a brief overview of the international human rights system is helpful. Human rights have at least four meanings. They can refer to normative principles about how humans are to be treated. They can stand for legal rights as such. More formally, they refer to the set of international human rights codified in human rights treaties or in other sources of international law. Finally, human rights are associated with the practice of institutions and actors that administer and enforce human rights.⁸

It should be noted at the outset that whether human rights should be the basis for international governance is contested. Human rights have been criticized for their Western origins and for their ineffectiveness. At a minimum, however, they provide an overarching vision for addressing issues of international concern, including certain AI applications. Virtually all countries are signatories to one or more of the major human rights conventions discussed immediately below and have agreed that the rights they establish are universal.⁹ There is a long history of their existence and of the institutions and practices that have emerged from them.¹⁰ At the international level human rights thus provide a common language and means to articulate and assess the positive and negative impacts of emerging technologies on people and societies.

At the international level, the primary set of formal rights is codified in treaties sponsored by the U.N., among them the Universal Declaration of Human Rights,¹¹ the International Covenant on Civil and Political Rights,¹² and the International

Covenant on Economic, Social and Cultural Rights.¹³ There are important regional human rights treaties: the African Charter on Human and Peoples' Rights,¹⁴ the American Convention on Human Rights,¹⁵ and, as mentioned, the European Convention for the Protection of Human Rights and Fundamental Freedoms.¹⁶ Human rights treaties are administered by organs contemplated by the treaties themselves or established for that purpose at the international and regional level.¹⁷ In the U.N. system, all U.N. bodies are supported by the Office of the High Commissioner for Human Rights.¹⁸ The regional treaties establish courts for dispute resolution: the Inter-American Court for Human Rights, the African Court on Human and People's Rights, and the European Court of Human Rights.

3.1 The U.N. Guiding Principles on Business and Human Rights: Respect for human rights and human rights due diligence and remediation

International human rights are a subset of international law and as such addresses business conduct only indirectly: in most cases, international law applies only to nation states and international organizations. States must first enact domestic legislation that applies international norms to companies under their jurisdiction, and several treaties that regulate business conduct do just that.¹⁹ In the alternative, states must consent to deep forms of regional integration. This is the case with the E.U., where regulations adopted at the E.U. level are automatically binding.

In human rights, business conduct has been governed by non-binding principles. Several documents create this framework,²⁰ but the U.N. Guiding Principles on Business

⁸ Nickel, J., 2021, "Human rights," The Stanford Encyclopedia of Philosophy, Fall 2021 ed., <https://tinyurl.com/yckbwa9j>

⁹ Vienna Declaration and Programme of Action, 1993, U.N. Doc. A/CONF 157/23

¹⁰ Latonero, M., 2018, "Governing artificial intelligence: upholding human rights and dignity," Data & Society, October 10, <https://tinyurl.com/3s8w2m33>

¹¹ Universal Declaration of Human Rights, 1948

¹² International Covenant on Civil and Political Rights, 1976

¹³ International Covenant on Economic, Social, and Cultural Rights, 1976. For a list of seven "core" international human rights instruments, see U.N. Population Fund, 2004, Core International Human Rights Instruments, <https://tinyurl.com/3kvfmurd>

¹⁴ African [Banjul] Charter on Human and People's Rights, 1981

¹⁵ American Convention on Human Rights, 1969

¹⁶ European Convention for the Protection of Human Rights and Fundamental Freedoms, 1950; article 19

¹⁷ For example, within the U.N. system, several bodies are formed under the U.N. Charter. These are the Human Rights Council, Universal Periodic Review, Special Procedures of the Human Rights Council, and the Human Rights Complaint Procedure.

¹⁸ U.N. Office of the High Commissioner for Human Rights, 2021 Human Rights Bodies, <https://tinyurl.com/59534hw6>

¹⁹ These include the OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions, the Paris Convention on the Third Party Liability in the Field of Nuclear Energy, the International Convention on Civil Liability for Oil Pollution Damage, the Council of Europe Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment, and the Hazardous Waste Convention. van den Herik, L., and J. Čerňič, 2010, "Regulating corporations under international law," *Journal of International Criminal Justice* 8:3, 725-743.

²⁰ According to Barnali Choudhury, these are (with current citations) the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct, 2023, OECD Publishing; the International Labor Organization Tripartite Declaration of Principles Concerning Multinational Enterprises and Social Policy, 2022, 6th ed., ILO Publishing; the UN Global Compact (based on corporate social responsibility principles); and the UN Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, 2011, U.N. Doc. A/HRC/17/31 [hereinafter Guiding Principles]. Choudhury, B., 2018, "Balancing soft and hard law for business and human rights," *British Institute of International and Comparative Law* 67:4, 961-986

and Human Rights have arguably been the most influential statement of the responsibilities of business in this area.²¹ The Guiding Principles establish three maxims: first, states have a responsibility to protect human rights; second, business firms should respect human rights; and third, victims should be given effective remedies for violations of those rights. As part of their responsibility to respect human rights, businesses “should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved.”²² The responsibility to address adverse human rights impacts involves not only a business’s own activities, but also seeking “to prevent or mitigate adverse impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.”²³

The Guiding Principles elaborate that management should adopt policies that commit to respect human rights,²⁴ but there are two more consequential requirements. First, there is the due diligence requirement: businesses should adopt a “human rights due diligence process to identify, prevent, and account for how they address their impacts on human rights.”²⁵ Such diligence will vary according to the size of the business and should be ongoing.²⁶ Due diligence encompasses both internal activities and business relationships: it should “cover adverse human rights impacts that businesses may cause or contribute to through [a business’s] own activities, or which may be directly linked to its operations, products or services by its business relationships.”²⁷ Businesses should then take “appropriate action” based on this human rights assessment.²⁸

Second, businesses are required to mitigate human rights harms. They should establish “[p]rocesses to enable the remediation of any adverse human rights impacts they cause or to which they contribute.”²⁹ States have primary responsibility for providing effective remedies for breaches of human rights, but “[w]here business enterprises identify that they have caused or contributed to adverse impacts, they should provide for or cooperate in their remediation through legitimate processes.”³⁰ This takes place mostly through “operational level grievance mechanisms” for adversely affected individuals or communities.³¹

3.2 The Guiding Principles as principles

Although the Guiding Principles set out in detail a business’s responsibilities regarding human rights, they are not legally binding. Barnali Choudhury notes that the principles were deliberately grounded in non-legal expectations and norms. There is no legal definition of “corporate responsibility to respect.” The principles do not impose any consequences for failing to meet these responsibilities, and there is no third-party oversight of compliance.³² However, despite their non-binding nature they have been highly influential. A working group established by the U.N. Human Rights Council to promote the principles claimed with justification that “[t]here is no doubt that the Guiding Principles have succeeded in providing a globally agreed-upon authoritative standard for what States and businesses need to do to respectively protect and respect the full range of human rights across all business contexts....”³³ They have been accepted by significant parts of the business community, including the large AI companies. Amazon, Apple, Google, IBM, Meta, and Microsoft have all stated that their human rights policies are informed in part by the Guiding Principles.³⁴

²¹ For example, the OECD Guidelines “draw from” the U.N. Guiding Principles, comment 41, and the ILO Tripartite Declaration states that the Guiding Principles “outline the respective duties and responsibilities of States and enterprises on human rights,” para. 10(a).

²² Guiding Principles Principle 11

²³ Id. Principle 13(b)

²⁴ Id. Principle 16

²⁵ Id. Principles 15(b), 17

²⁶ Id. Principles 17(b)-(c)

²⁷ Id. Principle 17(a)

²⁸ Id. Principle 19

²⁹ Id. Principle 15(c)

³⁰ Id. Principle 22

³¹ Id. Principle 29

³² Choudhury, *supra* note 20; pages 968-969

³³ U.N. Working Group on the issue of human rights and transnational corporations and other business enterprises, 2021, “Guiding Principles on Business and Human Rights at 10: taking stock of the first decade,” U.N. Doc. A/HRC/47/39, paragraph 11

³⁴ Amazon Global Human Rights Principles, <https://tinyurl.com/5nae36m5>; Apple, “Our commitment to human rights,” <https://tinyurl.com/yj7hs3kt>; Google, About Google: Human Rights, <https://tinyurl.com/2pjsw9um>; IBM, “IBM human rights statement of principles,” <https://tinyurl.com/3fwn45yh>; Sissons, M., 2021, Meta: “Our commitment to human rights,” <https://tinyurl.com/54jcccm9>; Microsoft: “Microsoft global human rights statement,” <https://tinyurl.com/mra5h7c7>

4. DUE DILIGENCE AS A LEGAL OBLIGATION

The adoption by corporations of the Guiding Principles can be viewed as consistent with the larger corporate social responsibility and environmental, social, and governance movements.³⁵ Over the past decade, some stakeholders have argued that the principles should be hardened into binding law, arguing that gaps in existing law allow for human rights abuses without recourse. The due diligence requirement has become the locus of these efforts and has become shorthand for a range of responsibilities set out in the Guiding Principles.³⁶ Some domestic legislation and regulations now require companies to engage in due diligence directed towards specific issues such as conflict minerals and forced and child labor.³⁷ Germany has adopted legislation that focuses on rights associated with labor and the environment.³⁸ France and Norway have been more expansive and have required larger companies to engage in human rights due diligence more generally.³⁹

This trend, however, is not without opposition. The push-pull is evident in the E.U. Artificial Intelligence Act; the proposed E.U. Directive on Corporate Sustainability; the Council of Europe Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law; and negotiations at the U.N. for a general treaty on business and human rights.

4.1 The E.U. Artificial Intelligence Act

By early 2024, the E.U. had completed most of the stages in approving the E.U. Artificial Intelligence Act (AIA), which had been under consideration since 2021. A final vote by the E.U. Parliament was expected in April 2024.

The general contours of the Act are well known, and a provisional text was released by the E.U. Council in late January.⁴⁰ When the law goes into effect, AI systems will be regulated in proportion to their risk of harm. AI systems that pose an unacceptable risk, such as those using subliminal techniques to distort a person's or group's behavior, are prohibited.⁴¹ Other systems are high-risk because they threaten "significant potential harm to health, safety, fundamental rights, environment, democracy, and the rule of law."⁴² Such high-risk AI systems are subject to a broad range of design, risk management, documentation, and reporting requirements.⁴³ General purpose AI models, such as large language models and other generative AI systems, are also regulated, particularly if they are declared to be "general purpose models with systemic risk."⁴⁴ AI systems that are not high-risk or general purpose models with systemic risk are subject to various transparency obligations.⁴⁵ For example, a company that uses a chatbot must disclose that an individual is interacting with an AI system.

By its terms, the AIA seeks to protect fundamental rights, particularly where prohibited AI practices or high-risk AI systems are concerned. National authorities that are empowered to protect those rights can gain access to documentation that companies who develop or deploy high-risk AI systems have submitted to regulators.⁴⁶ Further, all high-risk systems must have in place a risk management system, which among other things must identify and analyze known and reasonably foreseeable risks the high-risk system might pose to health, safety, or fundamental rights.⁴⁷ Providers of high-risk systems must give deployers instructions for use that include among other things information about known or foreseeable circumstances "which may lead to risks to health

³⁵ For a review of the literature on the effect of CSR and ESG governance measures on financial and stock performance, cost of capital, brand image and reputation, risk management and operational efficiency, and innovation, see Smit L., C. Bright, R. McCorquodale, M. Bauer, H. Deringer, D. Baeza-Breinbauer, F. Torres-Cortés, F. Alleweldt, S. Kara, C. Salinier, and H. Tejero Tobed, 2020, "Study on due diligence requirements through the supply chain, Final Report for the European Commission," 306-315, <https://tinyurl.com/2d6ztn9w>

³⁶ For a general discussion as of 2020, see *id.*, pages 192-212

³⁷ See, e.g., 17 C.F.R. § 240 13p-1 (United States, conflict materials); Child Labor Duty of Care Act, 2019 (Netherlands)

³⁸ Act on Corporate Due Diligence Obligations in Supply Chains, 2021 (Germany)

³⁹ Law n°2017-399 of 27 March 2017 Concerning the Duty of Vigilance of Parent Companies and Holding Companies (France); Bill for Responsible and Sustainable Business Conduct, 2021 (Netherlands); Act relating to enterprises' transparency and work on fundamental human rights and decent working conditions (Transparency Act) (Norway)

⁴⁰ European Parliament, 2023, "Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI," press release, December 12, <https://tinyurl.com/2wdpt9bk>. See also AIA Annex III. All references to the AIA are based on the provisional text released by the Council of the European Union, 2024 Interinstitutional File 2021/0106 (COD), January 26

⁴¹ AIA article 5

⁴² European Parliament *supra* note 40

⁴³ AIA articles 8-29, 51, 61-62

⁴⁴ *Id.* articles 52a-52d

⁴⁵ *Id.* articles 52, recitals 70-70e

⁴⁶ *Id.* article 64

⁴⁷ *Id.* article 9(2)(a)

and safety or fundamental rights.”⁴⁸ Such systems must be designed to allow meaningful human oversight aimed at preventing or minimizing risks to those rights.⁴⁹

The AIA also imposes a more formal human rights due diligence requirement for a limited number of AI systems. Public entities and private firms that provide public services, as well as operators deploying high-risk systems to evaluate a person’s creditworthiness or, in the area of insurance, to perform a risk assessment and to price life and health insurance policies, must perform a “fundamental rights impact assessment.”⁵⁰ The assessment must include a description of the deployer’s processes in which the high-risk system will be used, a description of the time and frequency of its use, categories of natural persons and groups likely to be affected, the specific risks of harm to such persons or groups, a description of the implementation of human oversight measures, and “the measures to be taken in case of the materialization of these risks, including their arrangements for internal governance and complaint mechanisms.”⁵¹ It is unclear from the language of the regulation whether mitigation processes and complaint mechanisms are mandatory, but it appears that there is a strong expectation that such mechanisms should be in place.⁵²

4.2 E.U. Directive on Corporate Sustainability Due Diligence

The E.U. has also been preparing the Directive on Corporate Sustainability Due Diligence. While the AIA marks an important step in regulating AI applications as such, if approved, the Due Diligence Directive will be a milestone in the hardening of international human rights norms for business in general.

Like the AIA, the final text of the legislation has not been officially approved as of this writing, but the Directive has been under consideration since 2022. The Directive was expected to receive final approval in March 2024 and a “final” draft has been circulated,⁵³ but late opposition by some member states has left in question the final details of the measure or whether it will be adopted at all.⁵⁴ As currently written, the legislation will affect all large firms,⁵⁵ including the major AI technology companies housed outside of the E.U.⁵⁶ The Directive largely codifies the U.N. Guiding Principles as further articulated by the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct, as well as international environmental norms. Member states must adopt legislation requiring large companies to conduct human rights and environmental due diligence.⁵⁷ This includes implementing due diligence policies and risk management systems, identifying and assessing actual and potential adverse human rights impacts, preventing and mitigating potential adverse impacts, bringing actual adverse impacts to an end and minimizing their extent, providing remediation, engaging with stakeholders, adopting a complaints procedure, monitoring due diligence policies and measures, and publicly disclosing its due diligence activities.⁵⁸

The Directive goes further and requires that businesses must be made subject to penalties for failure to meet the requirements of the Directive⁵⁹ and to civil liability – in the case of human rights, presumably to victims of human rights abuses.⁶⁰

⁴⁸ Id. articles 13(3)(b)(iii)

⁴⁹ Id. articles 14(2). Bias detection that requires the processing of personal data must be allowed “subject to appropriate safeguards for the fundamental rights and freedoms of natural persons.” AIA, art. 10(5)

⁵⁰ Id. article 29a(1); Annex III part 5. See also AIA recital 58g

⁵¹ Id. article 29a(1)

⁵² Id. recital 58g

⁵³ Directive on Corporate Sustainability Due Diligence Draft, 2024, January 24, <https://tinyurl.com/5n7bn26m> [hereinafter Due Diligence Directive]. All references are to the Jan. 24 draft.

⁵⁴ Wolters, L., 2024, press conference, Feb. 28, <https://tinyurl.com/4rvp4b7p> (announcing that a qualified majority in the E.U. Council needed to approve the legislation had not been achieved); Segal, M., 2024, “EU Council fails to approve new environmental, human rights sustainability due diligence law,” ESG Today, February 28, <https://tinyurl.com/5a58ppme>

⁵⁵ Companies with more than 500 employees and net worldwide turnover of €50 million. Due Diligence Directive, article 2(1)(a). Companies in the textile, agriculture and food processing, and mineral extraction and related industries are also covered if they have more than 250 employees and net turnover of €40 million. Id., article 2(1)(bb)

⁵⁶ Firms formed in third countries are subject to the Directive if they generated a net turnover of €150 million in the E.U., or had a net turnover of €40 million and operate in the sectors listed above. Id., article 2(2)

⁵⁷ Id., article 4

⁵⁸ Id., article 4(1)

⁵⁹ Id., article 20

⁶⁰ Id., article 22

4.3 Council of Europe Draft Convention on AI, human rights, democracy and the rule of law

The Council of Europe is also negotiating a framework convention on AI, human rights, democracy, and the rule of law. Meetings are planned in mid-March 2024 to finalize the text for submission to the Council of Ministers.⁶¹ The convention is intended to establish a “legal framework on the development, design, use and decommissioning of artificial intelligence, based on the Council of Europe’s standards on human rights, democracy, and the rule of law and other relevant international standards, and conducive to innovation...”⁶² At present, the major principles, rules and rights set out in the convention are organized into five areas: the application of AI systems by public authorities; the application of AI systems in the provision of goods, facilities, and services; fundamental principles of design, development, and application of AI systems; measures and safeguards for accountability and redress; and the assessment and mitigation of risks and adverse impacts. A primary issue still being negotiated is whether the treaty will apply to private entities. If so, signatories will be required to ensure that firms within their respective jurisdictions adhere to the terms of the treaty. The U.S., which is an observer in the negotiations, reportedly seeks to exclude private entities.⁶³ The E.U., on the other hand, supports applying the treaty to businesses.⁶⁴

4.4 Negotiations on a U.N. treaty on business and human rights

Work at the international level has been slower. Two years after the Guiding Principles were published, the U.N. Human Rights Council established an “open-ended intergovernmental working group” on transnational corporations and other business enterprises with respect to human rights, whose purpose is to elaborate an “international legally binding instrument to regulate in international human rights law, the activities of transnational corporations and other business enterprises.”⁶⁵

Work on the treaty has been ongoing for ten years, with no deadline for ending negotiations; thus, the final contours of the treaty are far from clear. As currently drafted, the treaty would require states parties to take measures to:

- a) prevent the involvement of business enterprises in human rights abuse;
- b) ensure respect by business enterprises for internationally recognized human rights and fundamental freedoms;
- c) ensure the practice of human rights due diligence by business enterprises;
- d) promote the active and meaningful participation of individuals and groups ... in the development and implementation of laws, policies and other measures to prevent the involvement of business enterprises in human rights abuse.⁶⁶

Like the other legislation already discussed, the treaty would also require states to enact legally enforceable obligations for businesses to engage in human rights due diligence.⁶⁷ States must also ensure that businesses take “appropriate steps to prevent human right abuses by third parties” when the business “controls, manages, or supervises” the third party.⁶⁸ The treaty has been opposed by the U.S. while the E.U. has pointed to its recent legislation to show that the E.U. is already taking action to ensure businesses respect human rights.⁶⁹

5. IMPLICATIONS FOR AI APPLICATIONS

If the efforts to require businesses to conduct human rights due diligence succeed, what are the implications for AI applications and the firms that develop, market, and use them? Under the E.U. AIA, which has all but been approved, firms that develop or use high-risk AI systems will be required to assess and to later mitigate the risks that such systems pose to fundamental rights. Because the formal

⁶¹ Council of Europe Committee on Artificial Intelligence, 2023, “Preliminary timeline for the negotiations,” CAI(2023)17rev2, December 11

⁶² Council of Europe, 2023, “Terms of reference of the Committee of Artificial Intelligence (CAI),” extract from CM(2023)131-addfinal, <https://tinyurl.com/4ddt6852>

⁶³ Bertuzzi, L., 2023, “EU’s AI ambitions at risk as US pushes to water down international treaty,” Euractiv, June 6, <https://tinyurl.com/5ybf743f>

⁶⁴ Bertuzzi, L., 2024, “EU prepares to push back on private sector carve-out from international AI treaty,” Euractiv, January 10, <https://tinyurl.com/v8w4j7rr>

⁶⁵ United Nations Human Rights Council, 2014, Res. 26/9, U.N. Doc. A/HRC/RES/26/9; page 2

⁶⁶ United Nations Human Rights Council open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights, 2023, updated draft legally binding instrument (clean version) to regulate, in international human rights law, the activities of transnational corporations and other business enterprises, <https://tinyurl.com/8pny3wnf>; article 6.2

⁶⁷ Id. article 6.4

⁶⁸ Id. article 6.5

⁶⁹ Annex to Emilio Rafael Izquierdo Miño (Chair-Rapporteur), 2021, Report on the seventh session of the open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights: Note by the Secretariat, U.N. Doc. A/HRC/49/65; page 24 (opening statement of the U.S.). The E.U. is not formally engaged in negotiations in the working group, but supports a legally binding treaty. OCHCR, 2023, “Note by the Secretariat: compilation of general statements from States and non-State stakeholders made during the ninth session of the open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights,” Office of the United Nations High Commissioner for Human Rights, <https://tinyurl.com/83hxmajk>

fundamental human rights assessment that applies to certain banking and insurance activities is intended to be “relatively easy to comply with,”⁷⁰ it can be argued that it should be equally straightforward to comply with the more general risk assessment and mitigation requirements that apply to high-risk systems. Companies will understandably pay closer attention to the technical and operational requirements the AIA imposes on those systems, but if only because of the financial costs of violating the AIA, firms that develop and use high-risk AI systems will nevertheless want to demonstrate that they have engaged in a fundamental rights assessment, particularly with respect to their business models. Studies of the risks posed by AI applications to human rights, such as those conducted by the E.U. Agency for Fundamental Rights and by B-Tech at the U.N. level discussed above could serve as starting points for that assessment.

If adopted, the Due Diligence Directive will be far more impactful if only for the large AI firms. There will be nuances as the individual member states implement its terms, but any due diligence requirement will almost certainly be much more detailed and, more importantly, will be grounds for liability if violated. In addition, it will involve monitoring parent, affiliated companies, and subsidiaries, as well as business partners (discussed below). Unlike with the AIA, the level of regulation will not vary with risk and the sources of international human rights law will be potentially broader. Virtually all the large AI companies already have established risk assessment and mitigation processes as part of their internal operations,⁷¹ but these will now be subject to regulatory assessment. As with the AIA, large companies developing or deploying AI should be aware of the human rights analyses conducted by the E.U. Agency for Fundamental Rights and secondarily by B-Tech, as well as by authorities within the member states where they are established or operate.

6. BROADER ISSUES

The recent evolutions in the E.U., the Council of Europe, and the U.N. mark a significant development in international human rights and business in general and AI applications in particular. They raise three closely related issues: 1) the extraterritorial application of regional and domestic law; 2) the tension between generality and specificity in international human rights law and its impact on the efficacy and feasibility of such law; and 3) the regulation of supply chains.

6.1 The extraterritorial application of regional and domestic law

When the U.N.’s work on a draft treaty began, John Ruggie, who led the work behind the Guiding Principles, did not oppose an overarching business and human rights treaty as such, but cautioned that several issues would need to be resolved for a treaty to represent true progress in advancing human rights. Among them, since states are already obligated to protect human rights, the next development in international law would involve requiring states to enforce their laws against companies for operations outside the territory. Ruggie observed that although some have argued that the extraterritorial application of human rights law is becoming a legal requirement, in his view, this was a step nation states were not willing to take.⁷² In this regard, when the Guiding Principles were adopted, a group of experts in international law and human rights issued the Maastricht Principles, concluding that current law does require states to protect human rights by enforcing them against the activities of its companies abroad.⁷³ However, not all observers agree with this contention,⁷⁴ and the law remains unsettled at this point.

Requiring states to enforce their laws abroad can be fraught. Under international law, a state can exercise jurisdiction over its own territory, actions abroad that have a direct effect in the

⁷⁰ AIA, Fundamental rights impact statement; pages 4-5

⁷¹ See, e.g., Microsoft Azure, 2022, “Foundations for assessing harm,” May 6, <https://tinyurl.com/5482xwex> (describing the company’s harms modeling approach to developing AI applications)

⁷² Ruggie, J., 2014, “A UN business and human rights treaty? An issues brief,” Harvard University Kennedy School of Government Business and Human Rights Resource Center, January 28, <https://tinyurl.com/3yzhxwff>

⁷³ Maastricht Principles on Extraterritorial Obligations of States in the Area of Economic, Social and Cultural Rights, 2013, art. 24; De Schutter, O., A. Eide, A. Khalfan, M. Orellana, M. Salomon, and I. Seiderman, 2012, “Commentary to the Maastricht Principles on Extraterritorial Obligations of States in the area of economic, social and cultural rights,” *Human Rights Quarterly* 34:4, 1084-1169

⁷⁴ Ruggie did not seem persuaded by this conclusion when he expressed his concerns about a general treaty in 2014. Ruggie supra note 72. See also Knox, J., 2011, “The Ruggie rules: applying human rights law to corporations,” in Mares, R., 2012, *The UN Guiding Principles on Business and Ethics: foundations and implementation*, Brill. Knox argues that the issue is not whether states can enforce human rights extraterritorially against those within its jurisdiction, but whether they are required to do so. Id. pages 78-79. He contends further that “[d]eveloped countries have generally opposed extraterritorial human rights obligations, and developing countries may not always like the idea, either, in the context of the duty to protect . . .” Id. page 82

territory, its citizens, or entities under its control, and universal crimes such as genocide. As discussed above, states can consent to greater integration, as is true with the E.U. By their terms, the AIA and Due Diligence Directive extend respectively to AI system providers and deployers and to large AI firms no matter where located, but both pieces of legislation justify this reach because they apply to actions by those entities in the E.U. or that have effects in it.⁷⁵ However, even though they are grounded in the standard rules for jurisdiction, E.U. legislation influences the business decisions of entities that fall under the jurisdiction of other states – this is the so-called Brussels Effect.⁷⁶ For some AI firms, European law is becoming the de facto regulator of AI applications.

To apply domestic law against a company abroad is to apply that law in a state that also has jurisdiction over that entity. States and regions adopt laws specific to their values and circumstances. The issue arises whether a “sending” state that applies its law abroad interferes with the domestic and regional policies of other states. The issue is not new: conflicts have arisen over states’ competition, anti-bribery, taxation, and discovery laws. With human rights, however, it can be argued that because states have agreed that they are universal, in theory it should not matter which state enforces them. But regions and states have differing views on the nature and scope of those rights; thus, the question arises how much discretion states should be given to articulate and enforce them. To grant too much leeway could weaken the universality of human rights, while to grant too little could lead to their being rejected as imposed from outside the state. The case can also be made under principles of subsidiarity that states should remain the primary locus for the regulation of firms. The legitimacy of international law is often questioned because such law is beyond the reach of ordinary citizens, whereas there are at least some mechanisms for public consensus at the domestic level. However, to remain at the status quo could result in human rights abuses being unremedied, which in turn has its own negative impact on the legitimacy of international law.

6.2 Specificity and generality in international Human Rights Law, efficacy, and feasibility

Like the question of discretion, in international governance there is a dilemma between on the one hand, crafting a treaty that has specific, enforceable norms only to have them rejected by some states, and on the other, drafting a more general treaty that garners greater participation but that is essentially toothless. Ruggie added that by its nature, business and human rights is an area comprising a constellation of laws and issues, so that any treaty will need to be written at a high level of abstraction. Only then can such a treaty encompass the entire field and garner the consent of states. Because of such generality, however, Ruggie feared that the resulting treaty would not be effective.⁷⁷

Either approach has positive or negative aspects for human rights. For example, the Council of Europe might conclude that for now it is better for countries such as the U.S. to participate in an AI human rights treaty that applies only to state actions and not to businesses. Thus the status quo would remain if, in contrast, the Council of Europe extends the treaty obligations to businesses, this would, of course, confirm a hardening of human rights norms for AI. But the country that houses most of the major AI companies would likely not be a party.

6.3 Supply chains and human rights

Public-facing companies, particularly the manufacturers of consumer goods, have long been asked to ensure that their suppliers conform to human rights standards. AI has not been immune from these efforts. Datasets labeled by people are still needed to train AI models. As is true with the garment industry, observers are concerned that data labeling is done offshore under adverse work conditions or that the people who label data will add their own biases in the labeling process.⁷⁸ AI companies have also been criticized for the downstream uses of their technology. This has been the case with facial recognition systems and other surveillance technologies.⁷⁹ The large technology companies have

⁷⁵ AIA recitals 24-25

⁷⁶ Bradford, A., 2012, “The Brussels effect,” *Northwestern University Law Review* 107:1, 1-67

⁷⁷ Ruggie supra note 72 page 3. He writes, “[T]he category of business and human rights . . . includes complex clusters of different bodies of national and international law—for starters, human rights law, labor law, anti-discrimination law, humanitarian law, investment law, trade law, consumer protection law, as well as corporate law and securities regulation.” *Id.*

⁷⁸ Tan, R., and R. Cabato, 2023, “Behind the AI boom, an army of overseas workers in ‘digital sweatshops,’” *Washington Post*, August 28, <https://tinyurl.com/58heee5r>; Rowe, N., 2023, “Underage workers are training AI: companies that provide Big Tech with AI data-labeling services are inadvertently hiring young teens to work on their platforms, often exposing them to traumatic content,” *Wired*, November 15, <https://tinyurl.com/pt8w54t8>; Springbord Blog, 2023, “The ethics of data labeling: ensuring fair and unbiased labeling,” June 20, <https://tinyurl.com/47nxpf6k>

⁷⁹ Weise, K., 2021, “Amazon indefinitely extends a moratorium on the police use of its facial recognition software,” *New York Times*, May 18, <https://tinyurl.com/mpbvrz52>

responded to these criticisms by adopting moratoria for certain uses, application processes, end user agreements, and terms of use that restrict the way in which these applications can be used.⁸⁰ But except perhaps in very limited circumstances, they are not considered liable for actions taken by those suppliers or customers.

In its current form, the proposed Due Diligence Directive and, to a lesser extent, the AIA, expand the responsibility of companies for partners in their supply chains. The Directive expressly “lays down rules ... on the obligations for companies regarding actual and potential human rights adverse impacts..., with respect to ... the operations carried out by their business partners in companies’ chains of activities[.]”⁸¹ For the most part, this would involve a company’s upstream activities, although the downstream disposal of products would be subject to the due diligence requirement as well. (For the time being, financial institutions would be exempt from taking into account its downstream business partners).⁸²

If the Directive is adopted, it would signal a major development in using human rights (and sustainability norms) to govern not only company activities, but also those of its partners. Recall that the Directive requires companies to among other things adopt due diligence policies and risk management systems, identify and assess actual and potential adverse human rights impacts, prevent and mitigate them, bring actual adverse impacts to an end and minimize their extent, and provide remediation. Several of these obligations require companies to involve themselves with the actions of their business associates. For example, as a company takes appropriate measures to prevent or mitigate potential adverse impacts, this includes among other things considering the impacts caused

by its business partners, taking into account the ability of a company to influence those partners.⁸³ Further, a company must seek contractual assurances from its direct business partners that they will follow the company’s code of conduct and in turn seek similar assurances from its partners.⁸⁴ (As discussed above, some technology companies already require this of their customers.) In extreme cases, a company could be required to cut ties with a business partner.⁸⁵

7. CONCLUSION

AI applications are capturing public attention just as human rights as a source of governance over business is evolving from a set of principles to legally binding obligations. This article has discussed efforts by policymakers to regulate AI through international human rights, surveying some of the human rights concerns that arise from AI applications, describing international human rights law as it applies to businesses in general, and reporting how those norms have been hardening at the domestic and regional levels. At the same time, the article has identified some of the issues that arise when international human rights are applied as legal obligations at the international level, particularly the problem of extraterritoriality and the dilemma of participation versus effectiveness in international agreements. However, because states have agreed that human rights are universal, they remain the appropriate framework for governing transformative technologies such as AI. Even though international actors will argue about the meaning and scope of these rights or whether specific AI applications even raise human rights concerns, no other framework provides better terms for vigorous debate and eventual consensus.

⁸⁰ For example, see Amazon.com, 2021, Notice of 2021 Annual Meeting of Shareholders and Proxy Statement, pp. 27–28 (describing some of Amazon’s controls over downstream uses of its technology)

⁸¹ Due Diligence Directive article 1(a)

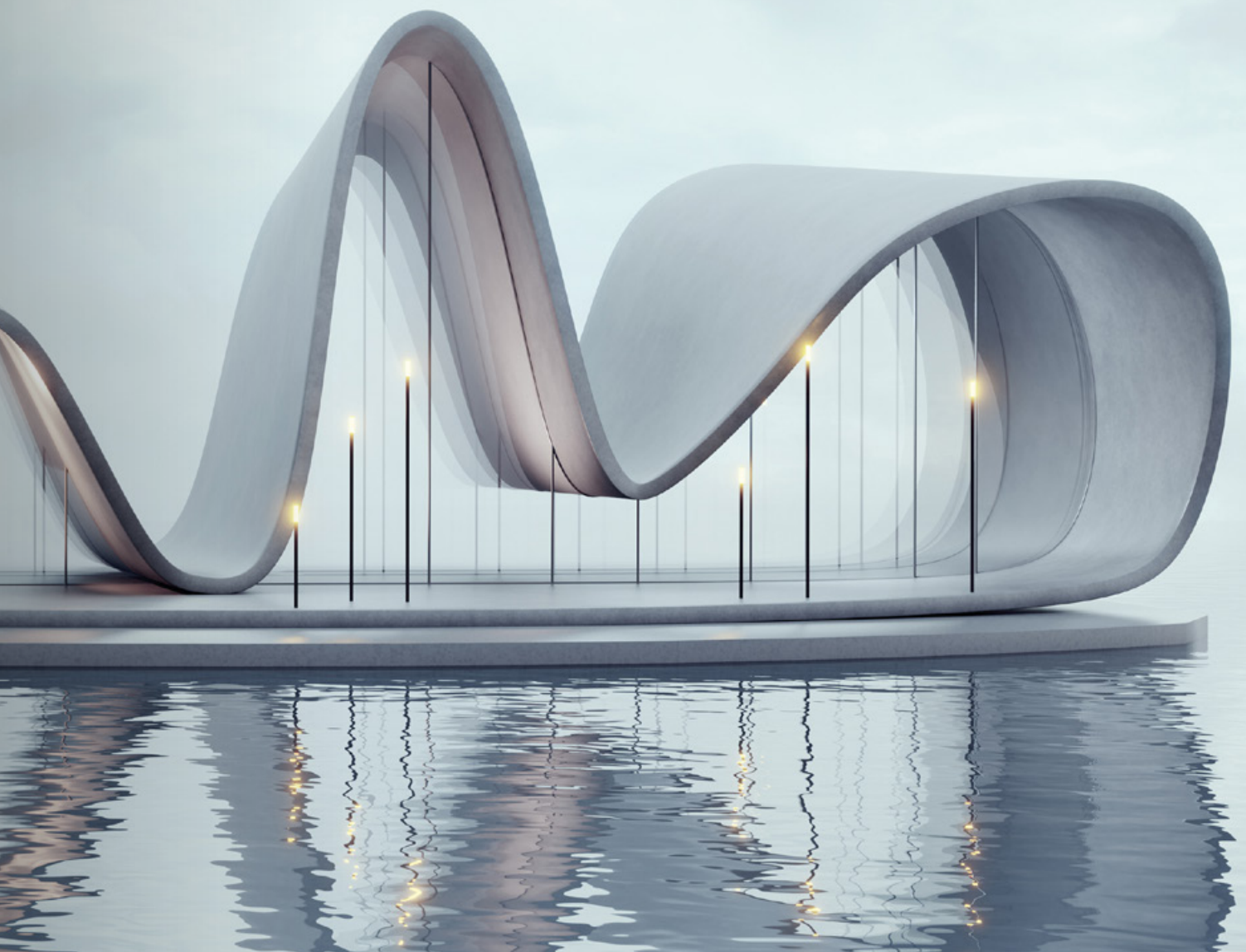
⁸² Id. recital 19

⁸³ Id. article 7(1)

⁸⁴ Id. article 7(2)(b). A direct business partner is defined in part as “an entity . . . with whom the company has a commercial agreement related to the operations, products or services of the company . . .” Id., art. 3(e)(i) An indirect partner is one that does not have such a commercial agreement, but which performs such services. Id. article 3(e)(ii)

⁸⁵ Id. article 8(6)

GOVERNANCE OF SUSTAINABILITY



82 Government incentives accelerating the shift to green energy

Ben Meng, Chairman, Asia Pacific, Franklin Templeton

Anne Simpson, Global Head of Sustainability, Franklin Templeton

92 Governance of sustainable finance

Adam William Chalmers, Senior Lecturer (Associate Professor) in Politics and International Relations, University of Edinburgh

Robyn Klingler-Vidra, Reader (Associate Professor) in Entrepreneurship and Sustainability, King's Business School

David Aikman, Professor of Finance and Director of the Qatar Centre for Global Banking and Finance, King's Business School

Karlygash Kuralbayeva, Senior Lecturer in Economics, School of Social Science and Public Policy, King's College London

Timothy Foreman, Research Scholar, International Institute for Applied Systems Analysis (IIASA)

102 The role of institutional investors in ESG: Diverging trends in U.S. and European corporate governance landscapes

Anne Lafarre, Associate Professor in Corporate Law and Corporate Governance, Tilburg Law School

112 How banks respond to climate transition risk

Brunella Bruno, Tenured Researcher, Finance Department and Baffi, Bocconi University

118 How financial sector leadership shapes sustainable finance as a transformative opportunity: The case of the Swiss Stewardship Code

Aurélia Fäh, Senior Sustainability Expert, Asset Management Association Switzerland (AMAS)

GOVERNMENT INCENTIVES ACCELERATING THE SHIFT TO GREEN ENERGY¹

BEN MENG | Chairman, Asia Pacific, Franklin Templeton
ANNE SIMPSON | Global Head of Sustainability, Franklin Templeton

ABSTRACT

Many government policies – both carrots and sticks – are driving the global transition to greener energy systems. In this article, we compare regulatory sticks, like carbon pricing, with carrots like feed-in tariffs that subsidized solar renewables in countries like Germany. We reviewed carbon pricing across the globe and discuss why higher prices remain challenging to implement politically. We also challenge the view that government subsidies are wasteful and discuss the steps taken by different countries to lower emissions. We conclude with an optimistic outlook of the U.S. government's new industrial policy and note a new record in global investments in low-carbon technologies. That said, governments in China, the E.U., and the U.S. are deploying carrots and sticks at markedly different speeds and intensity. Looking ahead, global security analysts seeking to generate alpha will need to integrate top-down subsidies into bottom-up security analysis to uncover risks and opportunities.

1. INTRODUCTION

Many government policies – both carrots and sticks – are driving the global transition to greener energy systems. In this article, we compare regulatory sticks, like carbon pricing, with carrots like feed-in tariffs that subsidized solar renewables in countries like Germany.

First, we review carbon pricing across the globe. Higher prices remain challenging to implement politically. We explain why some economists fixate on the efficiencies of carbon taxes and dismiss government subsidies as wasteful. We explore China's new carbon market, which aims to lower emissions from China's coal-fired power plants.

Second, we explain how governments like Germany helped kick-start a boom in solar-power innovations by deploying subsidized carrots. One of the biggest catalysts driving down today's solar prices comes from economies of scale in Chinese manufacturing. We review an emerging consensus among economists that subsidies are accelerating a “green vortex” in places like Texas in the U.S.

We conclude with an optimistic outlook of the U.S. government's new industrial policy and note a new record in global investments in low-carbon technologies. That said, governments in China, the E.U., and the U.S. are deploying carrots and sticks at markedly different speeds and intensity. Looking ahead, global security analysts seeking to generate alpha will need to integrate top-down subsidies into bottom-up security analysis to uncover risks and opportunities.

¹ This article draws inspiration from Bose, Dong, and Simpson (2019) and builds on the framework developed by Meng and Simpson (2023). The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of Franklin Templeton. This material is intended to be of general interest only and should not be construed as individual investment advice or a recommendation or solicitation to buy, sell or hold any security or to adopt any investment strategy. It does not constitute legal or tax advice. This material may not be reproduced, distributed or published without prior written permission from Franklin Templeton.

2. CARBON STICKS

For many years, the primary climate policy recommended by many economists was carbon pricing. Compared to government subsidies, carbon price signals offered a more elegant response to the complex problem of CO₂ emissions. Why? In their view, subsidies are often inflexible and inherently prone to wasteful overcapacity. With more countries racing to subsidize home-grown green industries, some warn that vast amounts of public money may go to waste [Economist (2023b)]. Instead of picking winners via government handouts – a “destructive new logic” that forsakes the invisible hand of free-market capitalism for the visible hand of “aggressive industrial policy” – carbon pricing offers a more efficient approach. Unlike subsidies, carbon pricing gives companies the freedom to reduce emissions by whatever means they see fit [Economist (2023c)].

If carbon pricing offers a more efficient road to our zero-carbon future, there is progress to celebrate. Over 46 countries price greenhouse gases – either through carbon taxes, emissions trading systems (ETS), or both – and they together account for 30% of global CO₂ emissions (Figure 1) [Black et al. (2022)]. One notable participant, China, launched the world’s largest

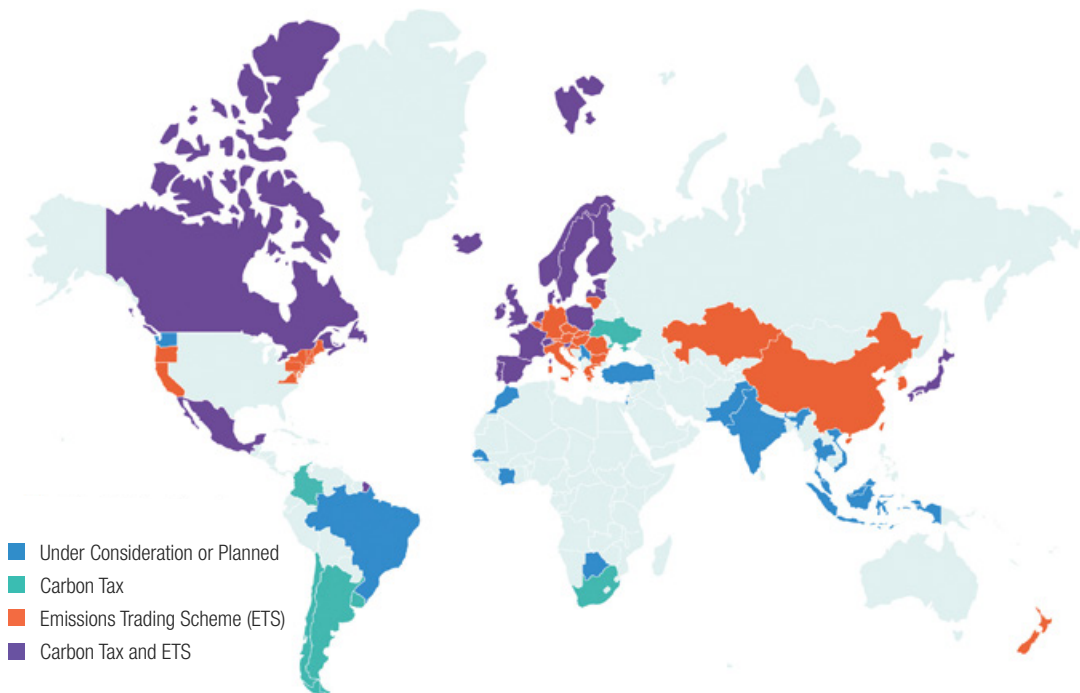
“

Today’s green vortex represents a handshake between the visible hand of government policies, which kick-start innovation with early funding, and the invisible hand of free-market capitalism, which efficiently directs capital to climate solutions.

”

carbon markets in 2021, covering one-seventh of global CO₂ emissions, and three times larger than the E.U.’s ETS [Busch (2022)]. Currently, China’s nation-wide ETS regulates roughly 2,162 companies from the country’s power generation sector, which emit 4.5 billion tons of CO₂ annually [Xue (2022)]. Given China is the world’s largest carbon emitter, we think this is a critical step in that country’s drive to reach zero carbon by 2060.

Figure 1: Countries choose different approaches to pricing carbon (as of August 2023)



Sources: World Bank Group (WBG), International Monetary Fund (IMF), and national sources.

Note: The boundaries and other information shown on any maps do not imply on the part of IMF any judgment on the legal status of any territory or any endorsement or acceptance of such boundaries.

At this early stage, China’s ETS is mainly structured to incentivize improvements at its coal-fired power plants by squeezing out inefficiencies and reducing carbon intensity [Mazzocco (2021b)]. China’s government initially planned to also include other high-carbon industrial sectors, such as cement and aluminum in 2022, but saw delays due to data quality. China’s Ministry of Ecology and Environment, for example, found compliance verification issues with most of the power sector company data [Tan (2022)]. By 2025, China aims to include even more carbon-emitting sectors, such as oil refining, chemicals, building materials, and non-ferrous metals. Looking ahead, India plans to launch its own national carbon market in 2026. Like China, India’s stakeholders will target high-carbon sectors such as power generation alongside a range of industrials like steel and cement [Choudhary and Macquarie (2023)]. Details of this cap-and-trade market – similar to the E.U.’s ETS – are still being worked out. For example, it is unclear how India’s existing voluntary carbon market will fit into the new trading scheme. That said, many of India’s stakeholders understand that carbon price signals need to be high enough that cutting emissions will be rewarded. To that end, India’s government plans to deploy a price stabilization mechanism to better incentivize low-carbon solutions [Singh and Narayan (2022)].

The framework for India’s pricing mechanism comes from the E.U., which added a carbon “market stability reserve” to its ETS in 2019. Just months after launching, E.U. carbon prices reached levels not seen in a decade [IEA (2020)]. Why? The supply of allowances had outstripped demand, causing a surplus. That meant carbon price signals were too low to incentivize economic changes. By tapping its reserve portfolio to buy carbon allowances, the E.U. has boosted carbon pricing to over U.S.\$100 per metric ton in 2022. As we discuss below, in the absence of stronger price signals, free markets can have difficulty reshaping economic activities.

Table 1: Carbon pricing via carbon taxes, emissions trading systems, or both

<p>Carbon taxes have a practical appeal by providing certainty over future emission prices that encourage green investments. These taxes also generate revenues that governments can use to tackle debt, ensure a more “just transition” by redirecting revenue to the poor and make green investments.</p>	<p>Emissions trading systems directly target emission levels by issuing carbon allowances that companies are required to obtain. By trading these allowances, the free market establishes carbon prices. It is not a fixed tax. Countries like France deploy fixed carbon taxes alongside the E.U.’s ETS.</p>
--	--

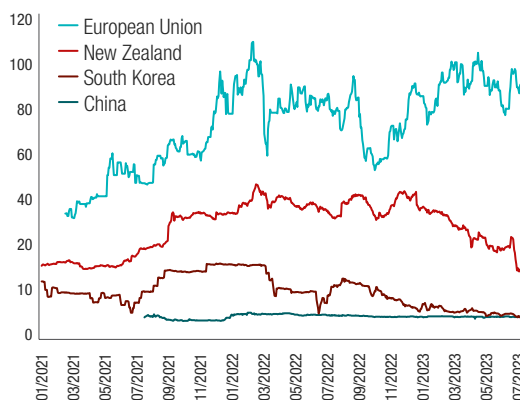
2.1 No pain, no gain

Since 2013, California’s ETS has had a clear mission. By setting limits for 85% of California’s CO₂ emissions, state authorities have established “a price signal needed to drive long-term investment in cleaner fuels and more efficient use of energy” [CARB (2015)]. In retrospect, however, a growing cohort of economists now admit these prices have not been tough enough to force much change on their own.

To be clear, California’s electric utilities have slashed emissions by 36% from 2013 through 2019 – but that was mainly due to state laws forcing utilities to incorporate more renewable power [Baker (2022)]. This critique is not unique to California. Back in 2012, economists reached the same conclusion when assessing Europe’s ETS. They found that the program had quite limited effects on the rate and direction of corporate clean-energy innovations [Schmidt et al. (2012)]. Thanks to the new price stability mechanism, however, the E.U.’s carbon price signals are exponentially higher today (Figure 2).

Two questions arise when looking at the global carbon-pricing map in Figure 1. First, how high are carbon prices today? Globally, the IMF estimates U.S.\$20 per ton on average across regions with price signals. Across all CO₂ emissions globally, however, it drops to U.S.\$5 per ton [Parry et al. (2022)]. In regions with price signals, only 10% have carbon prices at U.S.\$65 per ton or higher [OECD (2021)].

Figure 2: Emissions trading systems in the E.U., New Zealand, South Korea, and China (U.S.\$/metric ton CO₂ equivalent)



Source: Bloomberg (as of July 6, 2023)
 Note: Index currencies converted to U.S.\$, Korean Allowance Unit 2022: Listing on 1/4/2021 and delisting on 8/11/2023.

Second, how high should carbon price signals be? This depends on specific future goals: such as reaching net zero by 2050, calculating future carbon sequestration costs, or measuring the social costs of carbon (SCC) that each ton of carbon inflicts on humans. In 2013, an interagency working group within the U.S. government estimated that the SCC were U.S.\$36 per ton [Shelanski and Obstfeld (2015)]. Nine years later, new climate analysis by the U.S. Environmental Protection Agency raised the SCC to U.S.\$190 per ton [Lithgow (2022)]. This dovetails with 2022 economic research by Resources for the Future – a climate and energy think tank – that finds each additional ton of carbon emissions costs society U.S.\$185 [Rennert et al. (2022)].

It is worth noting here that the U.S. does not have a national ETS, nor do many other countries. Indeed, less than 30% of global CO₂ emissions are covered by carbon pricing schemes [IEA (2022)]. Out of this slice, the vast majority of today's CO₂ trading volume comes from just two carbon markets in the E.U. and China. Recent efforts to convince U.S. corporate CEOs and U.S. lawmakers to launch a similar ETS has come from the Commodity Futures Trading Commission (CFTC) [CFTC (2020)]. In testimony before the U.S. Senate in 2021, Bob Litterman, CFTC Climate-Related Market Risk Subcommittee of the Market Risk Advisory Committee Chairman, explained that without a national ETS, all manner of U.S. financial instruments – stocks, bonds, futures, bank loans – face painful and disorderly adjustments down the road [Litterman (2021)].

The CFTC's core message reflects the growing certainty that, outside the E.U., average carbon prices are simply too low to redirect capital at the scale and speed needed. Case in point, China's price is just U.S.\$8 per ton of CO₂, far below the E.U. (Figure 2). That said, we are less concerned for two reasons.

First, China's carbon pricing will reduce the carbon intensity of its coal-fired plants in the near term, before scaling up in the future. Second, the E.U. plans to implement a carbon border tax that will have positive ripple effects across the globe. Countries that trade regularly with the E.U. can either forfeit money at the border when selling high-carbon products or invest more at home in clean-energy systems to avoid the tax. We think the E.U.'s carbon stick will help incentivize trading partners to transition their economies quickly.

Indeed, in his Senate testimony, Litterman (2021) noted that the U.S. economy is 300% more carbon-efficient than competitors like China, Russia, and India. A carbon border adjustment would raise new revenues for the U.S. government. From Litterman's vantage, he said it was remarkable that leaders from both Republican and Democratic administrations have come together in support of a market mechanism that asks non-domestic manufacturers to compete based on carbon efficiency. "But given the win-win outcomes, it should not be surprising," he said.

2.2 Measuring carbon leakage

It is important to note that the E.U.'s carbon border adjustment mechanism (CBAM) remains a work in progress. For starters, the E.U. is initially targeting sectors it believes have the most significant risk of carbon leakage [E.C. (2023)]. That means high-carbon industrials, like iron and steel, aluminum, cement, fertilizers, as well as electricity and hydrogen. Many of these sectors, like cement, pose significant engineering and technology challenges, as we highlighted in 2021 [Khatoun et

Box 1: Spillover effects of a carbon border tax

By design, carbon border taxes are meant to have a global impact. But what about the spillover effects on emerging economies? Because many countries have either quite low or no carbon prices, some security analysts think companies outside the E.U. will simply shift their exports, like steel and fertilizer, to other non-E.U. countries and not bother decarbonizing [Sharma (2022)]. One think tank has modeled the cost increases that future E.U. carbon tariffs will have on iron and steel imported into the E.U. from China, Brazil, Russia, and India. Prices for India's steel could rise 15% in the E.U.; prices for steel from China, Brazil, and Russia could rise 3-4% [Xiaobei et al. (2022)]. The authors, however, note the macroeconomic impact of the border tax on these countries looks modest. For example, the effect on China's GDP is negligible – these exports into the E.U. are just 0.4% of China's overall exports – while Russia's GDP could drop 0.2% by 2030. Bear in mind, this economic analysis was published mere weeks after Russia's invasion of Ukraine.

al. (2021)]. Europe is deploying billions of capital in early-stage demonstration projects, testing green hydrogen and carbon capture solutions at steel and cement factories across Europe.

From now through the end of 2025, there will be no carbon tax at the E.U.'s borders. Instead, the focus will be on ironing out the methodology for accurately measuring the "Scope 1 emissions" embedded in these industrial goods. Scope 1 refers to direct CO₂ emissions during the production process. If nothing else, establishing the right methodologies to measure carbon, which is also verifiable globally, will be an enormous step forward.

These new methods are necessary to measure carbon leakage, which can happen in two ways. First, E.U. businesses could relocate industrial production to countries outside the E.U. with lower or no carbon prices. Second, carbon leakage can occur if products made in the E.U., like steel or cement, are replaced by equivalent imports with higher CO₂ intensity at cheaper prices.

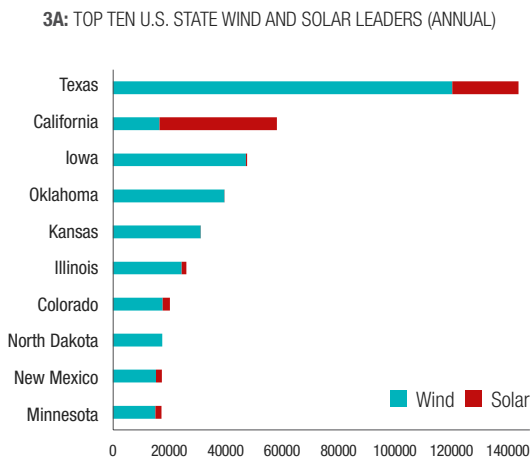
For security analysts, it is clear that E.U. carbon pricing brings headwinds to Europe's industrial companies. The cost of retrofitting plants with carbon capture, for example, are eating into profits and may boost prices higher than most non-E.U. competitors. Indeed, the "buy or sell" recommendations

of Europe's largest cement makers were downgraded in 2020 for this exact reason [Dempsey (2020)]. Analysts rightly argued that higher cement prices would expose E.U. companies to carbon leakage via cheaper imports from India's cement industry [Investec (2020)]. At the time, we noted a carbon border tax would likely resolve this issue. We stand by our analysis and think the macroeconomic impact on emerging economies will be modest – see our discussion on "spillover effects" in Box 1. We think Europe's border tax will lead the way to a faster energy transition across developed and emerging economies alike.

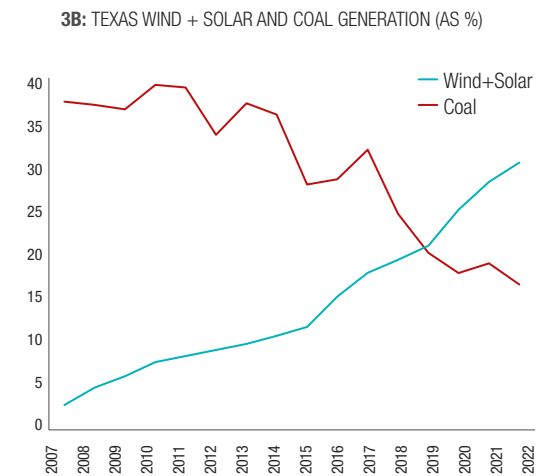
2.3 The green vortex

As we have discussed, carbon pricing has dominated conversations around climate policy for decades. Today, it still features prominently in academic circles and publications like *The Economist*. A growing number of scientists, however, now recognize that carbon sticks are not the only option. And they have clear evidence to prove it. Consider California's carbon market, which some climate analysts consider to be one of the best-designed carbon programs in the world [Hiltzik (2018)]. If that is true, how do we explain power generation in the state of Texas?

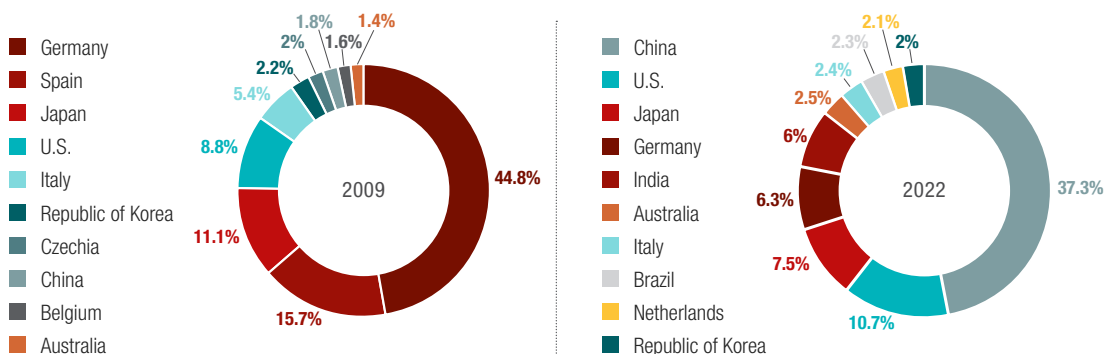
Figure 3: Texas' green vortex



Source: U.S. Energy Information Agency (EIA)
Net generation in thousand megawatthours as of 2022



Source: Electric Reliability Council of Texas (ERCOT)

Figure 4: Top ten countries by share of installed solar capacity (%)

Source: International Renewable Energy Agency (IRENA)

In the first quarter of 2022, Texas led the U.S. in renewable energy, accounting for over 14% of U.S. green-energy production [Gilligan (2022)]. Many Texans bristle at government taxes – the state does not levy a state income tax – and are proud of the state’s fossil-fuel industries. And yet, Texas now produces nearly twice as much electricity from renewables as from coal (Figure 3).

Texas is clearly decarbonizing. But why? Some climate analysts call this process a “green vortex” [Meyer (2021)]. The phrase describes the accelerating combo of technological advances and the appeal of green profits that were kickstarted by – wait for it – government subsidies. Today, we are seeing a newfound appreciation for industrial policy among economists, though certainly not all [Meckling (2021)]. This represents a qualitative shift away from classic climate policy that mainly focused on carbon pricing.

In our view, today’s green vortex represents a handshake between the visible hand of government policies, which kick-start innovation with early funding, and the invisible hand of free-market capitalism, which efficiently directs capital to climate solutions. All combined, the return premium from green climate solutions – a return “greenium” – is something we discuss in an upcoming paper in the *Journal of Investment Management*.

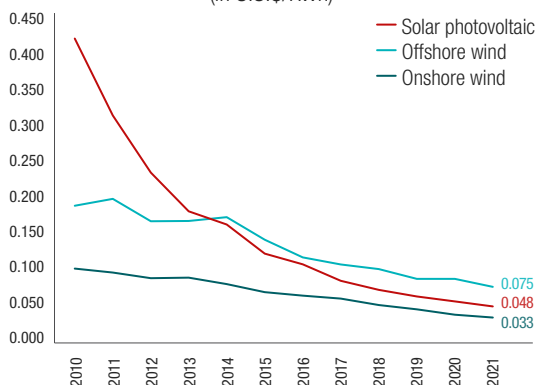
To unpack this worldview, we turn next to advancements in solar photovoltaic production in recent decades, which benefited from a wide range of government carrots such as loan guarantees and feed-in tariffs. Rather than imposing upfront costs on existing fossil-fuel assets, some policy analysts now argue clean-energy subsidies should precede phased-in taxes, to better redirect “private investment away from polluting capital and towards clean capital” [Rozenberg et al. (2020)].

3. SUBSIDIZED CARROTS

In October 2022, at the opening of the Chinese Communist Party’s 20th National Congress, President Xi Jinping spoke at considerable length about safeguarding the environment by accelerating China’s clean-energy revolution. To reach carbon neutrality by 2060, Xi reiterated the principle of “establishing the new before destroying the old” [Yin and Yep (2022)]. This phrase means building a reliable, renewables-centered economy first through government subsidies, before eliminating the use of fossil fuels like coal.

Xi’s philosophy is not unique to China. Researchers at the think tank MacroPolo remind us that advanced economies, chiefly Japan and Germany, deployed government loans and capital in the 1990s to help jump-start their fledgling solar industries. For example, Japan launched a solar rooftop subsidy program in 1994, helping drive down costs of solar installations by more than 65% over the following decade [Mazzocco (2021a)].

Figure 5: World's levelized cost of renewable electricity
(in U.S.\$/KWh)



Sources: Our World in Data, IRENA

Across Europe, but particularly in Germany, government feed-in tariffs were deployed as either a primary or exclusive policy mechanism to drive solar energy deployment through the 1990s and 2000s. Feed-in tariffs are government incentives that guarantee a certain level of financial benefit for each unit of electricity produced by renewables, like solar panels. These fixed-price contracts – which typically last 10 to 20 years – sent a clear price signal to developers and utilities across Europe that installing solar panels would be profitable [Couture et al. (2010)]. By substantially increasing these solar subsidies in 2000 and 2004, Germany saw an explosion of solar installations through the 2000s (Figure 4).

3.2 Green industrial policies

Around this same time, China was busy incentivizing solar panel manufacturing in rapidly urbanizing cities like Wuxi. China's manufacturers received access to subsidized land and modern manufacturing infrastructure, along with special financing and tax cuts. The goal was to accelerate growth in polysilicon manufacturing and wafer production, creating vertically integrated supply chains. The economist, Paul Krugman, calls this phenomenon, in which supplies of key materials, like polysilicon, are situated near the production of solar photovoltaic (PV) cells, modules, and panels, "agglomeration."

All combined, China's industrial carrots helped scale up solar PV production 500 times from 2000 to 2016 [Mazzocco (2021a)]. Why is scale important? Economists studying the mechanics of technological innovations find that economies of scale and learning-by-doing play an outsized role in lowering costs and improving quality across clean-energy technologies [Nagy et al. (2013)]. This economic theory – known as Moore's Law and, in a slightly modified version, called Wright's Law – was recently tested against historical data and held up quite well [Santa Fe Institute (2013)].

It is these economic laws – and the government incentives that drove them – that help to explain a seismic shift in competitiveness of renewable electricity over fossil fuel options. From 2010 to 2021, the costs of solar PV electricity dropped 88%, which is now below the costs of fossil fuel electricity (Figure 5) [IRENA (2022)]. At these prices, solar PV is now more profitable for power plants than coal- or gas-fired electricity.

This breakthrough in clean-energy pricing brings us back to the concept of the "green vortex" that we discussed earlier. In India, the outlines of a national carbon market are just emerging. And yet, it is with an eye toward green profits that India's largest power company is now committed to building 60 gigawatts of solar PV electricity by 2032 [Bullard (2021)]. Why? The power from newly built solar capacity in India is now cheaper than the power from existing Indian gas- and coal-fired power plants. It is really that simple. Indeed, India's government now plans to stop building new coal-fired power plants by removing a key clause from the final draft of its National Electricity Policy [Singh and Varadhan (2023)]. Cheaper renewables means India does not need new coal additions, apart from what is already in the near-term pipeline.

3.3 Leading with carrots

For investors worried that industrial policies may usher in the demise of free-market principles championed by Adam Smith, we highly recommend an economic paper from the Boston Review [Stokes and Mildenerger (2020)]. The authors have assembled a wide array of new research from economists who suggest government incentives – both industrial policy carrots and carbon pricing sticks – are indispensable to reaching our clean-energy future.

As for green-energy carrots overturning free-market orthodoxy, BloombergNEF (2021) notes that G20 governments handed out U.S.\$3.3 trillion of direct fossil-fuel subsidies from 2015 through 2019. These direct subsidies, however, do not include the mountain of implicit subsidies from governments that do not currently impose national carbon prices. The IMF recently calculated that governments showered companies with U.S.\$5.9 trillion of implicit fossil fuel subsidies in 2020 alone [Parry et al. (2021)]. If governments can hand out “carbon carrots” to oil and gas companies by avoiding an E.U.-style ETS, then subsidizing green-energy innovations should not scramble free markets, in our view.

As for solely focusing on carbon sticks to incentivize the energy transition, that approach can deliver short-term pain, like higher energy bills, while concealing longer-term gains for the environment, public health, and most economies. In our view, it is better to lead with government carrots that accelerate the arrival of cheaper green energy and well-paying jobs before phasing in higher carbon prices. In other words, we should build the new before destroying the old. This carrot approach has finally arrived in the U.S., first with infrastructure legislation in 2021, earmarking billions for a clean-energy grid

and charging stations for electric vehicles (EVs) [Newburger (2021)], and then with the Inflation Reduction Act (IRA) of 2022. The IRA offers U.S.\$369 billion in subsidies to jump-start clean-energy innovations while on-shoring green manufacturing [Hanwha (2022)].

These subsidies might be jarring to some security analysts. Some will point to Solyndra, a solar PV start-up that received a U.S.\$535 million loan guarantee from the U.S. government in 2009. In their view, Solyndra’s bankruptcy in 2011 is proof that government carrots are inherently wasteful. We note that Tesla received a similar loan for U.S.\$465 million in 2010 – part of the same program to accelerate U.S. clean-energy technologies – allowing it to expand its production facility [Bose et al. (2019)]. Was that loan also wasteful?

To understand how our security analysts scrutinize the impact of government carrots on capital markets and individual companies, we suggest reading an interview with our Shanghai-based investment team. They explain how integrating policies like “Made in China 2025” into equity and credit analysis helps uncover risks and opportunities that many investors might otherwise miss [Xu et al. (2021)].



4. CONCLUSION

If there is some handwringing over U.S. President Joe Biden's new industrial policies, The Economist notes that history offers some reasons for optimism. For example, in the aftermath of the Second World War, scores of governments unleashed industrial carrots to supercharge industrialization, with great success in places like Japan and South Korea [Economist (2023a)]. Today, the Biden administration is deploying similar incentives, like green-energy procurement contracts that will accelerate demand for 100 gigawatts of solar power systems over the next decade. That is nearly as much as the U.S.'s installed solar-power capacity today. It is an economic approach that harkens back to policies the U.S. deployed to land astronauts on the moon.

Responding to the U.S., the E.U. unveiled its own green industrial strategy in March 2023. While it does not offer new funding, the plan aims to simplify the thicket of E.U. regulatory

hurdles, streamlining the approval of national green-finance tools already available in Brussels [Economist (2023d)]. A major goal of building green industries inside the E.U. is reducing dependence on energy imports, a security lesson learned from Russia's war in Ukraine. The E.U. recognizes that China dominates global manufacturing across key net-zero technologies, including electric vehicle batteries, solar panels, and wind turbines [Campbell and Gritz (2023)].

So, what impact will these E.U. and U.S. industrial policies have? Over the long term, we see these programs expediting the push of green technologies forward, with competition between the world's three largest economies – the U.S., China, and the E.U. – reducing the costs of green technologies even faster [Conley (2023)]. Looking ahead, we believe the ability of investment analysts to produce alpha will increasingly hinge on analyzing how government carrots and sticks are accelerating both opportunities and risks across private and public investments.

REFERENCES

- Baker, D., 2022, "California's \$19 billion carbon market falls short in fight to curb emissions," Bloomberg, May 11, <http://tinyurl.com/yc7wrm4h>
- Black, S., I. Parry, and K. Zhunussova, 2022, "More countries are pricing carbon, but emissions are still too cheap," International Monetary Fund, July 21, <http://tinyurl.com/mry89vjc>
- BloombergNEF, 2021, "Climate policy factbook: three priority areas for climate action," Bloomberg Philanthropies, July 20
- Bose, S., G. Dong, and A. Simpson, 2019, The financial ecosystem: the role of finance in achieving sustainability, Palgrave Macmillan
- Bullard, N., 2021, "India's coal-dominated power market is tilting toward solar," Bloomberg, June 24, <http://tinyurl.com/52bbf5b7>
- Busch, C., 2022, "China's emissions trading system will be the world's biggest climate policy. Here's what comes next," Forbes, April 18, <http://tinyurl.com/3kye6k7r>
- Campbell, L., and A. Gritz, 2023, "Europe's green industrial policy and the United States' IRA," German Council on Foreign Relations, March 21, <http://tinyurl.com/5f9upcuh>
- CARB, 2015, "ARB Emissions Trading Program," California Air Resources Board, February 9, <http://tinyurl.com/2s65t7hn>
- CFTC, 2020, "Managing climate risk in the U.S. financial system," U.S. Commodity Futures Trading Commission, <http://tinyurl.com/3t6kzc9u>
- Choudhary, K., and R. Macquarie, 2023, "Disentangling India's new national carbon market," Sustainable Policy Institute Journal, Winter
- Conley, T., 2023, "Green subsidy race? 5 experts explain what to expect," World Economic Forum, March 30, <http://tinyurl.com/2zn3dzxx>
- Couture, T., K. Cory, C. Kreycik, and E. Williams, 2010, "A policymaker's guide to feed-in tariff policy design," National Renewable Energy Laboratory, July, <http://tinyurl.com/mv28dche>
- Dempsey, H., 2020, "Decarbonisation to drive 'dramatic' rise in cement prices, says Redburn," Financial Times, January 27, <http://tinyurl.com/5n8ykt56>
- E.C., 2023, "Carbon border adjustment mechanism," European Commission, <http://tinyurl.com/mu9w6adr>
- Economist, 2023a, "Warning from history for the new era of industrial policy," January 11, <http://tinyurl.com/yc6we8h8>
- Economist, 2023b, "Globalisation, already slowing, is suffering a new assault," January 12, <http://tinyurl.com/v6bw8cmp>
- Economist, 2023c, "The destructive new logic that threatens globalization," January 12, <http://tinyurl.com/2ynu44kb>
- Economist, 2023d, "What European business makes of the green-subsidy race," February 14, <http://tinyurl.com/576ta22x>
- Gilligan, C., 2022, "10 States that produce the most renewable energy," U.S. News & World Report, July 27, <http://tinyurl.com/yymdeevu>
- Hanwha, 2022, "As U.S. opens door to clean energy shift, Hanwha steps through," October 7, <http://tinyurl.com/3vv4aryy>
- Hiltzik, M., 2018, "Column: No longer termed a 'failure,' California's cap-and-trade program faces a new critique: Is it too successful?" Los Angeles Times, January 12, <http://tinyurl.com/bdft99k5>
- IEA, 2020, "Implementing effective emissions trading systems," International Energy Agency, July, <http://tinyurl.com/2xpf2mw2>
- IEA, 2022, "World energy outlook 2022," International Energy Agency, November, <http://tinyurl.com/42vbrt2s>
- Investec, 2020, "Global cement at a crossroads, India ahead on ESG," October, <http://tinyurl.com/4ubbuws9>
- IRENA, 2022, "Renewable power generation costs in 2021," International Renewable Energy Agency, July, <http://tinyurl.com/2mccctdas>
- Khatoun, B., A. Ness, L. Chow, P. Patel, and S. Ghosh, 2021, "Global innovations driving zero-carbon cement," Franklin Templeton, January 7, <http://tinyurl.com/5n76v9um>
- Litterman, R., 2021, "The cost of inaction on climate change," U.S. Senate Committee on the Budget, April 15, <http://tinyurl.com/4r3y9aa9>
- Lithgow, M., 2022, "US EPA proposed hiking social cost of carbon to nearly \$200/tonne," Carbon Pulse, November 11, <http://tinyurl.com/56wxxjyf>
- Mazzocco, I., 2021a, "Cheap solar (part 1): how globalization and government commercialized a fledgling industry," MacroPolo, January 14, <http://tinyurl.com/49pzryfu>
- Mazzocco, I., 2021b, "Beijing lines up the pieces for peaking emissions by 2030," MacroPolo, April 7, <http://tinyurl.com/3wsehsfw>
- Meckling, J., 2021, "Making industrial policy work for decarbonization," Global Environmental Politics, November 28, <http://tinyurl.com/3nspnssa>
- Meng, B., and A. Simpson, 2023, "Beyond ESG: government incentives delivering green transition," Carbonomics: the path to net zero, OMFIF Sustainable Policy Institute Journal, Winter, <http://tinyurl.com/mtfsfr6v>
- Meyer, R., 2021, "How the U.S. made progress on climate change without ever passing a bill," The Atlantic, June 16, <http://tinyurl.com/55npyttd>
- Nagy, B., J. Farmer, Q. Bui, and J. Trancik, 2013, "Statistical basis for predicting technological progress," PLOS One, February 28, <http://tinyurl.com/4err2wmm>
- Newburger, E., 2021, "Biden's infrastructure bill includes \$50 billion to fight climate change disasters," CNBC, November 15, <http://tinyurl.com/msx4vfep>
- OECD, 2021, "Carbon pricing in times of COVID-19: what has changed in G20 economies?" Organization for Economic Cooperation and Development, October 27, <http://tinyurl.com/vjpsysm>
- Parry, I., S. Black, N. Vernon, 2021, "Still not getting energy prices right: a global and country update of fossil fuel subsidies," IMF working paper no. 2021/236
- Parry, I., S. Black, and K. Zhunussova, 2022, "Carbon taxes or emissions trading systems?: Instrument choice and design," International Monetary Fund Staff Climate Notes, June, <http://tinyurl.com/5n6b63nk>
- Rennert, K., et al. 2022, "Comprehensive evidence implies a higher social cost of CO₂," Nature, October, <http://tinyurl.com/49hmubjp>
- Rozenberg, J., A. Vogt-Schilb, and S. Hallegatte, 2020, "Instrument choice and stranded assets in the transition to clean capital," Journal of Environmental Economics and Management 100:C, 102183
- Santa Fe Institute, 2013, "Study: 'Economy of scale laws' hold up well against observed data," Santa Fe Institute, March 5
- Schmidt, T., M. Schneider, K. Rogge, M. Schuetz, and V. Hoffmann, 2012, "The effects of climate policy on the rate and direction of innovation: a survey of the EU ETS and the electricity sector," Environmental Innovation and Societal Transactions, March, <http://tinyurl.com/y37je8fy>
- Sharma, M., 2022, "Europe's new carbon tariff won't help the climate," Bloomberg, December 21, <http://tinyurl.com/4kkrhf5n>
- Shelanski, H., and M. Obstfeld, 2015, "Estimating the benefits from carbon dioxide emissions reductions." The U.S. White House, July 2, <http://tinyurl.com/3mf7vh78>
- Singh, S., and M. Narayan, 2022, "Exclusive: India to bolster carbon trading market with stabilisation fund," Reuters, December 21, <http://tinyurl.com/aujtrhpa>
- Singh, S., and S. Varadhan, 2023, "Exclusive: India amends power policy draft to halt new coal-fired capacity," Reuters, May 4, <http://tinyurl.com/2xezfhdb>
- Stokes, L., and M. Mildenerger, 2020, "The trouble with carbon pricing," Boston Review, September 24, <http://tinyurl.com/4ahr3n2a>
- Tan, L., 2022, "One year in: China's national emission trading system," Refinitiv, July 25, <http://tinyurl.com/4cbaba7p>
- Xiaobei, H., Z. Fan, and M. Jun, 2022, "The global impact of a carbon border adjustment mechanism: a quantitative assessment," Task Force on Climate, Development and the International Monetary Fund, March,
- Xu, L., T. Liu, and W. Fei, 2021, "Local knowledge is the key to China credit analysis," Franklin Templeton, June 28, <http://tinyurl.com/4jhn4uyt>
- Xue, Y., 2022, "China's national carbon trading scheme marks one-year anniversary, with analysts expecting stricter regulation and data monitoring ahead," South China Morning Post, July 16, <http://tinyurl.com/r5zb3vae>
- Yin, I., and E. Yep, 2022, "China to boost development of new energy system, climate change governance: Xi," S&P Global Commodity Insights, October 17, <http://tinyurl.com/2urzsd4er>

GOVERNANCE OF SUSTAINABLE FINANCE

ADAM WILLIAM CHALMERS | Senior Lecturer (Associate Professor) in Politics and International Relations, University of Edinburgh

ROBYN KLINGLER-VIDRA | Reader (Associate Professor) in Entrepreneurship and Sustainability, King's Business School

DAVID AIKMAN | Professor of Finance and Director of the Qatar Centre for Global Banking and Finance, King's Business School

KARLYGASH KURALBAYEVA | Senior Lecturer in Economics, School of Social Science and Public Policy, King's College London

TIMOTHY FOREMAN | Research Scholar, International Institute for Applied Systems Analysis (IIASA)

ABSTRACT

This article offers insights into what sustainable finance means and how it is addressed in the public policy context using a subset of the Carrots & Sticks dataset that comprises 1,070 sustainable finance policies. The study reveals the financial services sectors targeted, who is governing, and how binding sustainable finance policies are. Additionally, the study explores whether policymakers and standard-setters concentrate their efforts on recommending positive action or establishing binding rules. The findings help to advance a shared understanding of the governance of sustainable finance in the context of public policymaking.

1. INTRODUCTION

There is an increasing expectation that public policy can incentivize the financial services industry toward sustainable activities and assets, and away from ones that harm people and the planet. One of the key points of focus of such regulatory efforts is around disclosure requirements. The rationale being that requiring greater reporting will bolster transparency into financial holdings, and this can instigate market pressures away from financing “brown” assets, and towards greener activities.

Despite widespread interest in the topic, there is a paucity of knowledge of how sustainable finance is being governed. Dimmelmeier (2021) offers insight into how sustainable finance has evolved as a “contested concept” since the late 1990s. Kumar et al. (2022), in their large-scale review of the state of the art in academic literature, find that the definition of sustainable finance remains broad, “encompassing myriad dimensions of sustainable ways to attain finance and investment goals.” Indeed, sustainability remains an issue

area characterized by opacity. The Economist (2022) named “sustainability” one of the “woolliest words in business” and the 2022 World Economic Forum meetings in Davos were preempted with articles that strove to detail “what is sustainable finance and how it is changing the world” [Broom (2022)].

In this article, we advance our understanding of the governance of sustainable finance. We do this by using natural language processing (NLP) techniques to analyze the 1,070 policies in the Carrots & Sticks database¹ that focus on sustainable finance. We reveal which activities (e.g., asset management, banking, and insurance) are targeted and how binding policies are. Our goal is to mitigate opacity and the persistence of terms as merely a “North Star” [i.e., loosely defined principles, see van den Broek and Klingler-Vidra (2021)], offering clarity by detailing how sustainable finance is conceived and operationalized in public policy. In addition, we assess whether the aim of sustainable finance policy is “hard” law or “soft” law [Abbott and Snidal (2000)], using “carrots” or “sticks”, to effect action.

¹ Carrots & Sticks (carrotsandsticks.net)

Our findings help us to advance a shared understanding of the governance of sustainable finance in terms of whether Copenhagen, Glasgow, or Rio – a metaphor for this broader suite of public policies – are sufficiently targeted, ambitious, and sector-focused. This has several important and direct policy implications as a lack of conceptual clarity around sustainable finance can create confusion among stakeholders and the general public, lead to inconsistent and ineffective policy outcomes, result in implementation challenges, and can reduce accountability as it applies to policy success or, perhaps more importantly, failure.

2. SUSTAINABLE FINANCE: A 30-YEAR ODYSSEY OF A CONCEPT

What is sustainable finance? A quick answer would be that it depends on who you ask and when. Different and competing definitions have evolved over time and as a reaction to changing policy exigencies [Schoenmaker (2017), Schoenmaker and Schramade (2019)]. To an important degree, sustainable finance is a “contested concept” [Dimmelmeier (2021)] replete with enough ambiguity to encompass “myriad dimensions of sustainable ways to attain finance and investment goals” [Kumar et al. (2022)].

As evidenced in recent stocktaking exercises of 227 articles in Bui et al. (2020), 166 articles in Cunha (2021), and 936 articles in Kumar et al. (2022), research on sustainable finance is vast. At the same time, however, there is no consensus on the meaning of what we label “sustainable finance”. In a recent overview, Forstater and Zhang (2016) explain how, instead of a single definition, there are “a few working definitions and sets of criteria.” Policymakers, practitioners, and academics use different terms to refer to the same thing.

This includes a broad range of related but different neologisms. For the European Commission and the United Nations Global Compact, the preferred term is “sustainable finance”. However, the OECD and the International Financial Corporation (IFC), as well as governments in the U.K., Germany, and China, use “green finance” and “green banking”. We also see the use of terms like “climate finance” (World Bank), “(socially) responsible investing” (Principles for Responsible Investing (PRI)); Code for Responsible Investment in South Africa (CRISA), and “sustainable investing” (Global Sustainable Investment Alliance). The United Nations Environmental

Program (UNEP), tracing the term back to its origins with the 1992 Earth Summit in Rio de Janeiro,² uses “sustainable finance” and “green banking”. While some of these terms have distinct and well-defined meanings, they are often used to refer to the same broad concept. Recent academic stock-taking exercises, including Dimmelmeier (2021), Kumar et al. (2021), and Akomea-Frimong et al. (2021), confirm the same use of a broad range of different terms in the academic literature. Policy institutes, including the Stockholm Sustainable Finance Centre³ and Swiss Sustainable Finance⁴, while noting the absence of a common terminology, propose sustainable finance lexicons with no less than 100 entries.

Underlying this conceptual confusion, however, is a unifying feature of sustainable finance – namely the core idea of how finance (both investing and lending) interacts with economic, social, and environmental issues [Schoenmaker and Schramade (2019), Köbel et al. (2020), Kumar et al. (2022), Lindenberg (2014), Urban and Wojcik (2019)]. Bakken (2021) defines it as investing in line with environmental, social, and governance (ESG) considerations. Rather than only an “E” or green focus, researchers assert that sustainable finance refers to the ways by which finance (both investing and lending) interacts with ESG issues [Schoenmaker and Schramade (2019), Kumar et al. (2022), Urban and Wojcik (2019)].

But what does this mean in the world of governance? The first tack policymakers take amounts to making general declarations about leveraging finance toward sustainability ends. For the OECD, “green finance” is defined as “achieving economic growth while reducing pollution and greenhouse gases”.⁵ For the IFC (2009), green finance is defined as “[i]nvestment products that preserve the environment, ensure social justice and promote economic prosperity.” A second approach is to shift focus to how funds are channeled by investors. This is described in terms of “investments flowing to sustainable development projects” (International Development Finance Club), “resources” catalyzing climate resilient development (World Bank), as well as “capital rising for projects with environmental benefits (Green Bonds Principles). A third tack places emphasis on the investor. The idea here is about getting ESG information to investors and ensuring they “consider” ESG factors when making investment decisions [European Commission; Code for Responsible Investment in South Africa (CRISA), the Global Sustainable Investment Alliance]. The United Nations’ PRI provides the

² <http://tinyurl.com/8vupkve6>

³ The Stockholm Sustainable Finance Centre’s sustainable finance lexicon can be found here: <http://tinyurl.com/yc4ucx5j>

⁴ The Swiss Sustainable Finance glossary can be found here: <http://tinyurl.com/4m9632ae>

⁵ OECD Green Finance and Investment, <http://tinyurl.com/96cxssfu>

“

Our findings stress the importance of balancing regional and international policies with strengthened national transparency requirements across all ESG pillars. ”

clearest expression of this idea: responsible investing is about “explicitly acknowledging the relevance to the investor of ESG factors and the long-term health and stability of the market as a whole.”

To help cut through the confusion of sustainable finance as it is presented in governance contexts, we use natural language processing (NLP) techniques to examine the current operationalization of sustainable finance as an issue area, and the nature of policies globally in terms of their binding nature.

3. DATA AND METHODS

Our analysis uses data from Carrots & Sticks (C&S), an online database and policy repository of corporate sustainability policy. C&S takes a broad approach to defining corporate sustainability policy and, as of 2023, comprised 2,463 policy instruments in 132 countries, 76 international and regional organizations, in 39 languages, and ranging from 1897 to present [Chalmers et al. (2024)]. C&S acts as a platform of platforms, bringing together and consolidating information from other databases including European Corporate Governance Institute (ECGI), Green Policy Platform, PRI, the Reporting Exchange (RE), and the Sustainable Stock Exchange Initiative (SSE), each of which aggregates corporate sustainability policies, broadly conceived.⁶

Using C&S’s corpus of sustainability policy documents, we identified all policies that specifically target financial activities and institutions. To do this, and building on the work of Al-Ubaydli and McLaughlin (2017) and state-of-the-art natural language processing (NLP) techniques more generally [Rice and Zorn (2021), Gentzkow et al. (2019), Loughran and McDonald (2016)], we first established a bespoke dictionary of “n-grams”, or unique terms [including both single words or unigrams, as well as terms with two words (bi-grams), and three words (tri-grams)] that refer to four distinct sets of financial activities: (1) banking, (2) financial market infrastructure (FMI), a category that includes securities and commodity exchanges, (3) fund management, and (4) insurance.⁷ The four categories were created by combining the codes and descriptors of two widely used schemes for classifying distinct sectors of economic activities, namely: the United Nations’ International Standard Industrial Classification scheme (ISIC rev. 4); and the North American Industry Classification System (NAICS).⁸ Through an iterative process of careful hand coding amongst the five members of the research team, we generated a set of unique n-grams for each of these four distinct financial activities.⁹ Our n-grams are meant to be categorical, exhaustive, and allow for deviations in spelling, pluralization, and punctuation. Relative to previous studies, this allows us to assess “who in finance” is targeted by sustainable finance policies.

The result is a corpus of 1,070 sustainable finance policies (i.e., corporate sustainability policies that target financial activities and institutions) spanning a time period from 2001 to 2021. Given our focus on the specificities of language and linguistic change over time, only English language policies were retained. The corpus includes policies from 95 countries from all major world regions as well as 23 international organizations. We analyze this corpus of sustainable finance policies through a combination of hand-coding and NLP techniques.

⁶ Corporate sustainability includes “corporate responsibility”, “corporate social responsibility”, “environment, social, governance”, “ESG”, “materiality”, “non-financial materiality”, “shared value”, and “social value”. It does not include the broader suite of labor-related governance policies, such as “industrial relations”, “labor reforms”, and “labor regulation”.

⁷ Banking consists of commercial banking, savings institutions, credit unions, and other depository credit intermediation, as well as securities and contracts brokerages. FMI refers to many of the services auxiliary to banking and financial market activities, such as securities and commodity exchanges, loan brokers, financial transactions processors, and investment advisors. Fund management includes pension funds, health and welfare funds, open-end investment funds, and portfolio management. Insurance encompasses life, health, medical, property, and casualty insurance carriers, as well as claims adjusting, reinsurance, and insurance brokers.

⁸ This approach to developing finance ngrams builds on the work of Al-Ubaydli and McLaughlin (2017). A key difference, however, is that these authors treat finance as a single category, conflating everything from commercial banking to fund management to pensions and central banking. Our approach is more nuanced and allows us to not only distinguish between various sub-sectors of financial activity (i.e., banking, fund management, insurance, etc.), but it also allows us to pinpoint specific activities in each sector. For instance, in the banking category, we can distinguish between specific banking activities (like stock broking, commercial banking, and credit granting) and institutions (like savings banks and commercial banks) and financial instruments (like money orders and unlisted equities).

⁹ The ngram dictionary can be found at <http://tinyurl.com/3ds4z7bx>

4. ANALYSIS AND RESULTS

4.1 What terminology is being used?

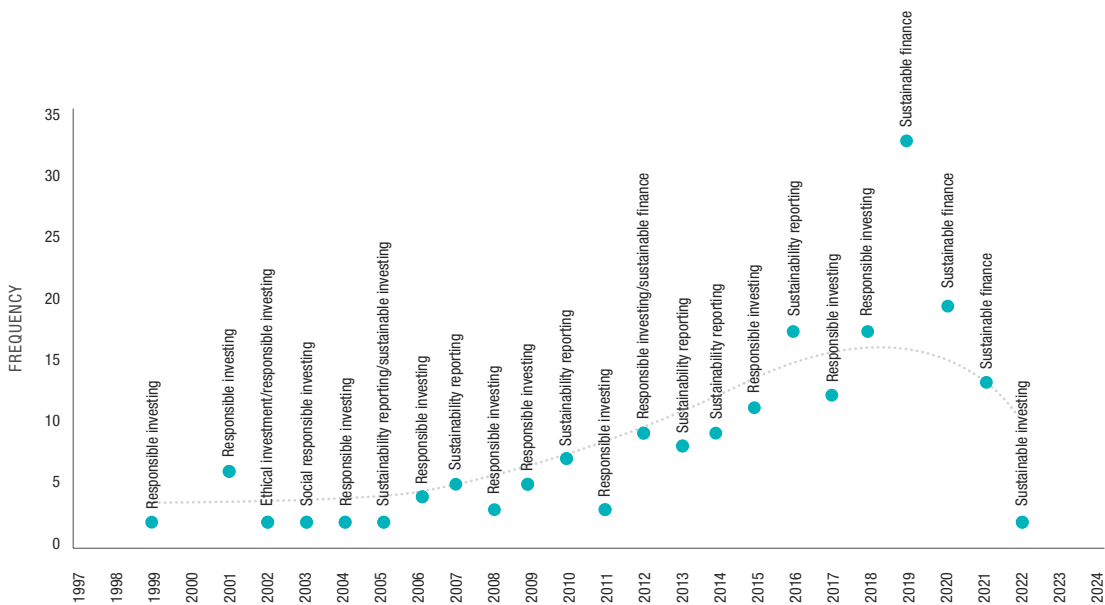
The language of sustainable finance, as this term is now understood, dates from around the beginning of the 21st century. The first appearance of a sustainable finance n-gram in our corpus is in 2001. In this year, the European Union’s “Green paper promoting a European framework for corporate social responsibility” uses the terms “responsible investing”, “socially responsible investing”, and “ethical investing”. In the same year in the Netherlands, Stichting Corporate Governance Onderzoek Pensioenfondsen’s Corporate Governance Handbook, uses “responsible investing” and “socially responsible investing” as cognates. This timeline aligns well with broader developments in the field. In particular, only a few years before, in 1997, the term “sustainable development” first appeared in the Delphi and Ecologic Institute’s report on “The role of financial institutions in sustainable development.” According to Dimmelmeier (2021), just two years later in 1999 “the term ‘Responsible Investment’ appeared for the first time in the continental mainstream news” [see also, Gond and Boxenbaum (2013)].

Over time, we see an increase in the relative prevalence of n-grams involving “green”, “carbon”, and “climate” (“E” language), with green bond, green finance, green investing, and climate finance all rising in frequency over time [Chalmers et al. (2023)]. At the same time, other than “responsible investing”, the language of social responsibility (“S” language) dropped from its previously prominent position. The prominence of “E” n-grams in 2016-2021 relative to previous time periods suggests the term sustainable finance has gone full circle since its genesis in Rio in 1992. While it originated from an environmental perspective, the language of sustainability appears to have been consumed by that of social responsibility in the early 21st century. We are now seeing the proliferation of “climate” and “green” in public policies pertaining to instruments (e.g., green bonds) for acting on the “E” aspect of sustainable finance.

Figure 1 shows the most frequently used sustainable finance n-grams for each year between 1998 and 2022, as well as its frequency of use. We see that only a few n-grams dominate the sustainable finance landscape. In fact, from a total of 40 possible n-grams, Figure 1 only includes five sustainable finance n-grams.

Figure 1 only captures the most frequently used n-gram for each year of our analysis. Are we seeing the use of broader array of different SF n-grams (over time)?

Figure 1: Most frequently used sustainable finance n-grams (2001-2021)



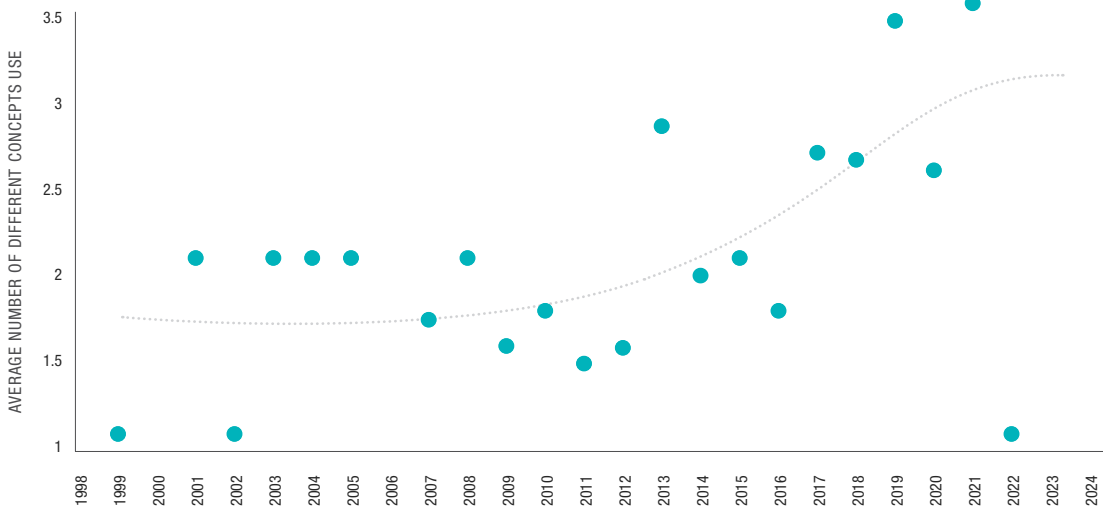
Note: The gray dashed line represents a LOWESS trend, i.e., the smoothed relationship between the data points.

In contrast, Figure 2 shows average number of different SF n-grams used per document across the same period. The general trend is that sustainable finance policies are using an increasing number of different SF concepts over time. By 2015, we see an average of nearly three distinct n-grams in use per policy and about 3.5 by 2021. These averages do not convey the more extreme diversity of terms used in certain outlier policies. Take the Sustainable Stock Exchange Initiative's (SSE) 2017 "How stock exchanges can grow green finance: a voluntary action plan" as an example. In this single policy document, the SSE uses a total of 12 different sustainable finance n-grams, including sustainable finance instruments (carbon tax, green bonds, and green securities) as well as cognate terms (sustainable economy, green investing, sustainable investing, responsible investing, sustainability reporting, climate finance, sustainable finance, green finance, and blended finance).

Using multiple different concepts can foster a lack of conceptual clarity and could contribute to concept stretching. This, in turn, can make it difficult to develop effective governance, leading to, or exacerbating:

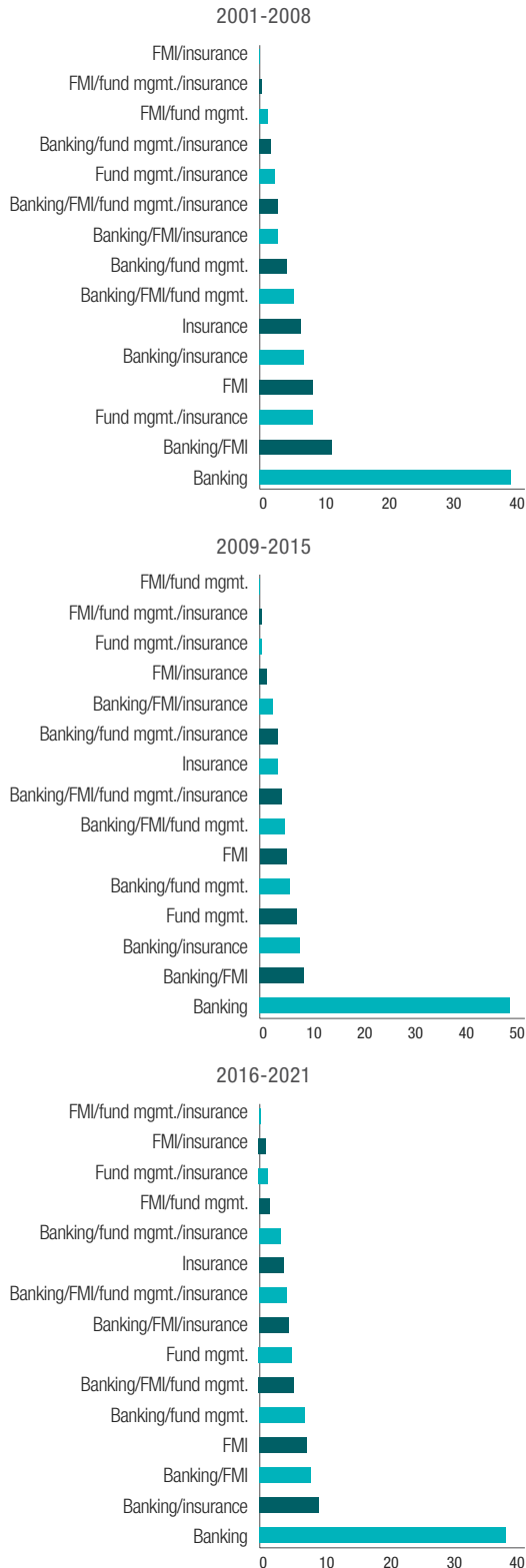
- **Inconsistent policy outcomes:** policies developed using different concepts may not address the same issues, leading to inconsistent results.
- **Implementation challenges:** if policymakers use different concepts to describe the same issue, it can make it difficult to develop a coherent policy framework that can be effectively implemented.
- **Reduced accountability:** if the concepts used to describe an issue are unclear or inconsistent, it can be challenging to assess whether the policy has been successful or not.
- **Missed opportunities:** if policymakers use different concepts to describe the same issue, they may miss important aspects of the problem, leading to incomplete or ineffective policy solutions.

Figure 2: Average use of different sustainable finance n-grams within a single policy document



Notes: The gray dashed line represents a LOWESS trend, i.e., the smoothed relationship between the data points.

Figure 3: Share of sustainable finance policy by financial sector target over time



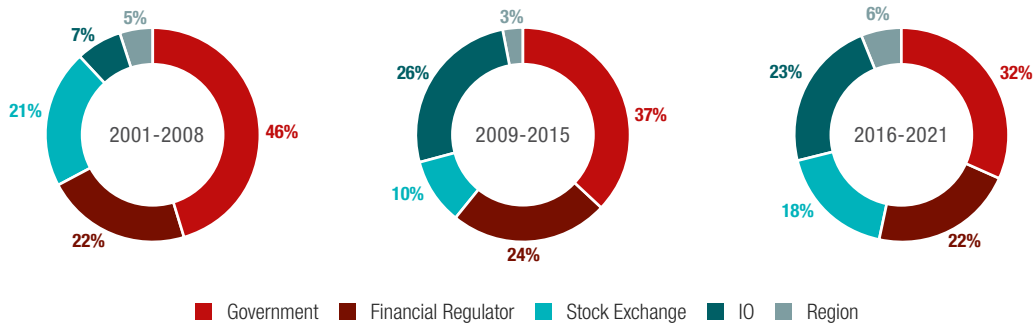
4.2 Which financial sectors are targeted in governance, and who is making policy?

While the language of sustainable finance policy is important, so too is understanding which parts of the financial system are being targeted by such policies and which organizations are most active in sustainable finance policymaking.

Addressing the first of these issues, Figure 3 details the proportion of policies within our sustainable finance corpus targeting the different sub-sectors of finance, namely: banks, financial market infrastructure (FMI), fund management, and insurance. To better trace these developments over time, we examine trends in three time periods. The first period starts with the first appearance of a sustainable finance cognate term in public policy 2001 and runs until 2008; the second is from 2009-2015; and the third is from 2016-2021. These time periods roughly align with the recent stocktaking exercises of Kumar et al. (2022) and Dimmelmeier (2021). Important here is our ability to capture sufficiently long time periods and to isolate key events that likely shaped sustainable finance policy, like the 2008 global financial crisis, CoP 2009, and the 2015 Paris Climate Agreement. As some policies reference more than one sector at a time, we also include combinations of sectors targeted. Banking is by far the most targeted sector in all three time periods. By contrast, policies with a sole focus on fund management, FMI, and insurance are far less common, with each of these sectors having roughly similar shares. There has been relatively little movement in these sector shares over the past two decades. While the prominence of banks within our sustainable finance policy corpus is unsurprising, the scale of this focus perhaps is. According to the Financial Stability Board [FSB (2022)], in 2021, banks accounted for 37.6% of global financial assets yet around three-quarters of the documents in our corpus reference banks in some capacity.

Which institutions are more active in issuing these policies? As illustrated in Figure 4, national governments are responsible for the largest proportion of the policies in our corpus. This dominant role of national governments has declined over time, however, falling from 46% of issuing all policy documents at the beginning of this century to 32% in the latest period. This gap has been filled by international organizations (IOs), whose share has increased from just 7% in the first period to 23% at the end. This likely mirrors the establishment of new sustainable-finance focused IOs, like the SSE and PRI, as well as increased focus on sustainable finance by the likes of the U.N. Global Compact and the OECD. In contrast to IOs, regions, like the European Union, consistently feature very little as issuers of sustainable finance policy. Finally, stock exchanges and financial regulators account for around

Figure 4: Sustainable finance policy by issuer type



20% of policies each. This rise of IOs as sustainable finance policy issuers augurs well for future cross-border and multi-stakeholder collaborations. What remains to be seen, though, is how successful these policies are, or promise to be, in driving towards substantive actions.

4.3 How binding are sustainable finance policies?

Researchers distinguish between policies that are binding and enshrined in legislation or “hard law”, and all other types of non-binding or “soft law” policies [Abbott and Snidal (2000)]. Where do sustainable finance policies fall on this spectrum, and to what extent are different issuer types (e.g., national governments and IOs) writing hard or soft laws? To investigate this, we use an existing dictionary of 183 “constraining” or restrictiveness terms specifically developed to analyze legal, legislative, and regulatory documents [Loughran and McDonald (2011)]. This dictionary includes terms related to degrees of commitments, compulsion, dictates, mandates, and obligations.

The results of this analysis are presented in Figure 5, which shows the mean of the count of restriction n-grams per issuer type (i.e., IOs, national governments, etc.) per year in our sustainable finance policy corpus. Higher scores correspond with a greater degree of “restrictiveness”. As a baseline to put our results in context, we include a restrictiveness score for Basel III (10.3%), the set of international banking standards developed in response to the Global Financial Crisis.

Figure 5 paints a mixed picture of more binding policies in some sectors and less binding policies in others, with no clear overall time trend. First, financial regulators and IOs tend to issue relatively unrestrictive sustainable finance policy guidelines and there has been little change in this approach over time. Though Copenhagen, Glasgow, and Rio have convening power, our findings align with the idea that

Figure 5: Average policy restrictiveness by issuer type over time



Notes: The gray solid line represents a LOWESS trend, i.e., the smoothed relationship between the data points. The blue dots are average restrictiveness scores for all policies issued in that year.

IOs continue to issue “soft law”, rather than policies that have teeth in requiring action. Second, perhaps surprisingly, national governments have moved substantially over the past decade towards issuing “softer” policies. The policy with the single highest restrictiveness score in our corpus is the 2001 Australian Financial Services Reform Act, which mandated all issuers of financial products to disclose the extent to which “labor standards or environmental, social or ethical considerations are taken into account in the selection, retention, and realization of an investment”. Since then, national governments have sought instead to encourage action through establishing best practices, codes, and strategies rather than mandating or requiring actions or disclosures. Finally, stock exchanges and regional actors, like the European Union, have become somewhat more restrictive in their policymaking in recent years, although signals here are quite noisy with large changes year-to-year. Stock exchanges have moved closer towards issuing “hard law” policies – this raises some concerns about the scope of companies being targeted by such policies (e.g., only publicly-traded firms).

5. DISCUSSION AND CONCLUSION

Our research identifies substantial evolution in the nature and scope of sustainable finance governance over time. One trend is the increasing emphasis placed on the environment and climate finance. Terms such as “green bonds” and “green investing” now feature prominently. This shift reflects a growing recognition of the finance sector’s pivotal role in mobilizing large amounts of private capital to meet investment needs for achieving the U.N. SDGs and the climate targets of the Paris Agreement [UNEP (2015), Bielenberg et al. (2016)].

A second trend – the increasing share of sustainable finance policies issued by international organizations – reflects the increased emphasis on multilateral efforts in recent years, such as the Glasgow Financial Alliance for Net Zero and the Taskforce on Climate-Related Financial Disclosures. And third, the lack of any clear direction in the restrictiveness or bindingness of these policies (i.e., the degree to which they are “hard” or “soft” law) raises potential governance concerns [Abbott and Snidal (2000)]. Notably, policies from regional entities, such as the E.U., and by stock exchanges have become more restrictive over time, while national government policies



have become less restrictive. This divergence underscores the nuanced governance landscape, prompting concerns about the efficacy of high-profile international collaborations like the Glasgow convention in achieving meaningful results without follow-up enforcement by national governments.

The emergence of less restrictive national policies seems consistent with the hopes being placed on carbon offset markets, which are currently nascent.¹⁰ Offsets are traded on voluntary carbon markets (VCMs), which operate outside of regulatory purview and allow (but do not require) companies to invest in a variety of emissions-reducing activities, including renewable energy, agricultural, and forestry-related projects. The voluntary nature of the VCMs raises concerns about the quality and validity of the purchased offsets. For example, the emissions reduced or sequestered must be additional to those under a business-as-usual scenario and must be both verifiable and persistent – challenges that can be especially difficult to meet for forestry and other land-use projects. Recently, blockchain-based initiatives in the forestry sector have aimed to address these concerns, though the value of these initiatives are yet to be demonstrated. And while still in a pilot or proposal phase, partnerships between corporations and project developers are emerging [Kotsialou et al. (2022)]. There are concerns, however, that household demand for such green finance is still muted [Bethlendi et al. (2022)].

In stark contrast, carbon allowances traded in compliance markets, such as the E.U.'s Emission Trading Scheme (ETS) portray maturity, with an estimated value of more than U.S.\$100 bln in 2020 [Blaufelder et al. (2021)]. Historically, the effectiveness of these schemes has been blunted by low implied carbon prices. However, recently the price of permits, specifically in the E.U.'s Emission Trading Scheme (ETS), has increased dramatically and is expected to grow further, as the cap on emissions will continue to decrease annually until 2030. With E.U. ETS being the flagship program for achieving ambitious climate targets, these dynamics align with our research findings, indicating a trend towards increased restrictiveness in policies implemented by regional organizations such as the E.U. over time.

In conclusion, our research highlights a notable rise in sustainable finance policies and terminology since 2001, with a growing emphasis on environmental factors. However, there is a concerning trend of decreasing restrictiveness in national government policies, countered by stricter measures from regional entities like the E.U. and stock exchanges. Market incentives are gaining prominence over concrete obligations for companies, while global events like Copenhagen, Glasgow, and Rio provide guiding principles. Nonetheless, our findings stress the importance of balancing regional and international policies with strengthened national transparency requirements across all ESG pillars.

¹⁰ The estimated market value of the carbon offsets market was around £300m in 2020 [Blaufelder et al. (2021)].

REFERENCES

- Abbott, K. W., and D. Snidal, 2000, "Hard and soft law in international governance," *International Organization* 54:3, 421-456
- Akomea-Frimpong, I., D. Adeabah, D. Oforu, and E. J. Tenakwah, 2021, "A review of studies on green finance of banks, research gaps and future directions," *Journal of Sustainable Finance & Investment* 12:4, 1241-1264
- Al-Ubaydli, O., and P. McLaughlin, 2017, "RegData: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997-2012," *Regulation & Governance* 11, 109-123
- Bakken, R., 2021, "What is sustainable finance and why is it important?" Harvard Extension School blog, August 9, <http://tinyurl.com/36s2r3m2>
- Bethlendi, A., L. Nagy, and A. Pora, 2022, "Green finance: the neglected consumer demand," *Journal of Sustainable Finance & Investment*, <http://tinyurl.com/rwn5pyfp>
- Bielenberg, A., M. Kerlin, J. Oppenheim, and M. Roberts, 2016, "Financing change: how to mobilize private-sector financing for sustainable infrastructure," McKinsey Center for Business and Environment, <http://tinyurl.com/49dcesfw>
- Blaufelder, C., C. Levy, P. Mannion, and D. Pinner, 2021, "A blueprint for scaling voluntary carbon markets to meet the climate challenge," McKinsey & Co., January 29, <http://tinyurl.com/55d9d7x5>
- Broom, D., 2022, "What is sustainable finance and how is it changing the world?" World Economic Forum, January 20, <http://tinyurl.com/3dfn6fj2>
- Bui, T. D., M. H. Ali, F. M. Tsai, M. Iranmanesh, M. Tseng, and M. K. Lim, 2020, "Challenges and trends in sustainable corporate finance: a bibliometric systematic review," *Journal of Risk and Financial Management* 13:11, 1-26
- Carvalho, B., A. Wiek, and B. Ness, 2021, "Can B Corp certification anchor sustainability in SMEs?" *Corporate Social Responsibility and Environmental Management*, <http://tinyurl.com/56rc46mz>
- Chalmers, A. W., R. Klingler-Vidra, and O. van den Broek, 2024, "From diffusion to diffuse-ability: a text-as-data approach to explaining the global diffusion of Corporate Sustainability Policy," *International Studies Quarterly* 68:1, <http://tinyurl.com/33njvbk>
- Chalmers, A. W., R. Klingler-Vidra, D. Aikman, K. Kuralbayeva, and T. Foreman, 2023, "Sustainable finance: the rise of the 'E' in ESG," *Global Policy*, February 9, <http://tinyurl.com/y96y3hux>
- Climate Bonds Initiative, 2022, <http://tinyurl.com/yfzyatzc>
- Cunha, F. A. F. S., E. Meira, and R. J. Orsato, 2021, "Sustainable finance and investment: review and research agenda," *Business Strategy and the Environment* 30:8, 3821-3838
- Dimmelmeier, A., 2021, "Sustainable finance as a contest concept: tracing the evolution of five frames between 1998 and 2018," *Journal of Sustainable Finance & Investment* 13:4, 1600-1623
- FSB, 2022, "Global monitoring report on non-bank financial intermediation 2022," *Financial Stability Board*, December 20, <http://tinyurl.com/ydpdx6a>
- Forstater, M., and N. N. Zhang, 2016, "Background note: definitions and concepts," *UNEP Inquiry Working Paper* 16/13
- Gentzkow, M., B. Kelley, and M. Taddy, 2019, "Text as data," *Journal of Economic Literature* 57:3, 535-574
- Gond, J.P., and E. Boxenbaum, 2013, "The glocalization of responsible investment: contextualization work in France and Québec," *Journal of Business Ethics* 115, 707-721
- IFC, 2009, "Financing a sustainable future," *International Finance Corporation*, World Bank Group, <https://tinyurl.com/3ttuhv7n>
- Köbel, J. F., F. Heeb, and T. Busch, 2020, "Can sustainable investing save the world? Reviewing the mechanisms of investor impact," *Organization & Environment* 22:4, 554-574
- Kotsialou, G., K. Kuralbayeva, and T. Laing, 2022, "Blockchain's potential in forest offsets, the voluntary carbon markets and REDD+," forthcoming in *Perspectives at Environmental Conservation*
- Kumar, S., D. Sharma, S. Roa, W. M. Lim, and S. K. Magla, 2022, "Past, present and future of sustainable finance: insights from big data through machine learning of scholarly research," *Annals of Operations Research*, <http://tinyurl.com/3jevyp8k>
- Lindenberg, N., 2014, "Definition of green finance," *German Development Institute*, <https://tinyurl.com/2kxs2yzw>
- Loughran, T., and B. McDonald, 2011, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *Journal of Finance* 66:1, 35-65
- Loughran, T., and B. McDonald, 2016, "Textual analysis in accounting and finance: a survey," *Journal of Accounting Research* 54:4, 1187-1230
- Maltais, A., and B. Nykvist, 2020, "Understanding the role of green bonds in advancing sustainable finance," *Journal of Sustainable Finance & Investment*, <http://tinyurl.com/mr46p7tz>
- Rice, D. R., and C. Zorn, 2021, "Corpus-based dictionaries for sentiment analysis of specialized vocabularies," *Political Science Research and Methods* 9, 20-35
- Schoenmaker, D., 2017, "Investing for the common good: a sustainable finance framework," *Bruegel*
- Schoenmaker, D., and W. Schramade, 2019, *Principles of sustainable finance*, Oxford University Press
- Shishlov, I., R. Morel, and I. Cochran, 2016, "Beyond transparency: unlocking the full potential of green bonds," *Institute for Climate Economics Report*, <http://tinyurl.com/5crr476m>
- SSE, 2017, "How stock exchanges can grow green finance: a voluntary action plan," *Sustainable Stock Exchange Initiative*, <https://tinyurl.com/4d6au8mb>
- The Economist, 2022, "The woolliest words in business," May 14, <http://tinyurl.com/3tps44h7>
- U.N., 2021, *IPCC Climate Change Report*, <http://tinyurl.com/yb4xmyxk>
- UNEP, 2015, "The financial system we need: aligning the financial system with sustainable development," *United Nations Environment Programme*, <http://tinyurl.com/yrrn6kn>
- UNEP, 2017, "The evolution of sustainable finance," June 6, <http://tinyurl.com/8vupkve6>
- Urban, M. A., and D. Wojcik, 2019, "Dirty banking: probing the gap in sustainable finance," *Sustainability* 11, 1-23
- van den Broek, O., and R. Klingler-Vidra, 2021, "The UN Sustainable Development Goals as a north star: how an intermediary network makes, takes, and retrofits the meaning of the Sustainable Development Goals," *Regulation & Governance* 16:1, <http://tinyurl.com/47nycaw7>
- Wijen, F., and M. E. Flowers, 2022, "Issue opacity and sustainability standard effectiveness," *Regulation & Governance*, <http://tinyurl.com/4kz5mexr>
- World Bank, 2019, "10 years of green bonds: creating the blueprint for sustainability across capital markets," *World Bank*, March 18, <http://tinyurl.com/99r6xmre>

THE ROLE OF INSTITUTIONAL INVESTORS IN ESG: DIVERGING TRENDS IN U.S. AND EUROPEAN CORPORATE GOVERNANCE LANDSCAPES

ANNE LAFARRE | Associate Professor in Corporate Law and Corporate Governance, Tilburg Law School

ABSTRACT

This article explores the divergent regulatory, political, and societal trends in Europe and the U.S. regarding the environmental, social, and governance (ESG) rights and duties of institutional investors. While the SEC in the U.S. has demonstrated a greater focus on stricter ESG disclosure rules, political debates persist, reducing ESG discussions to mere ideology. In contrast, Europe exhibits a significant surge in sustainable finance and corporate governance, emphasizing transparency obligations outlined in regulatory initiatives like the SFDR. Examining the tools available to institutional investors, this article delves into the disparities in duties imposed on them in the U.S. and Europe and scrutinizes the voice tools they employ for promoting ESG goals as active owners, with a particular focus on shareholder sustainability proposals. In conclusion, this article highlights the need for a more harmonized and effective approach to sustainable investment. It advocates aligning European aspirations for sustainable capital allocation in the member states with increased emphasis on sustainability voice, potentially through a forthcoming new Shareholder Rights Directive (SRD III).

1. INTRODUCTION

In a period marked by major global concerns over sustainability challenges, greater attention has been paid to responsible business and financial practices. Institutional investors are facing pressure to actively use their influence regarding environmental, social, and governance (ESG) issues within the companies they choose to invest in.¹ In theory, these investors can have a central role in steering environmentally friendly corporate behaviors, including encompassing endeavors to diminish carbon emissions.² They possess the ability to direct funds towards sustainable investments and hold significant shareholder rights and engagement tools. These range from informal shareholder interactions like meetings, calls,

and letters to actively voting against managerial proposals.³ But in practice, there are some important doubts about their actual role in the transition towards more sustainable business activities.

Several institutional investors have openly expressed their commitment to corporate sustainability through different channels, including the yearly Letter to CEOs from BlackRock CEO Larry Fink. However, the actual impact and depth of their engagement remains debatable, with research emphasizing concerns about greenwashing practices. Some authors question the sustainability preferences of investors, especially “The Big Three” (BlackRock, Vanguard, and State Street Global Advisors), raising doubts about genuine commitment

¹ For instance, Strine, L., 2019, “Toward fair and sustainable capitalism: a comprehensive proposal to help American workers, restore fair gainsharing between employees and shareholders, and increase American competitiveness by reorienting our corporate governance system toward sustainable long-term growth and encouraging investments in America’s future,” University of Pennsylvania, Institute for Law & Economics research paper no. 19-39.

² Ringe, W-G., 2021, “Investor-led sustainability in corporate governance,” ECGI Law Working Paper 615/2021

³ McCahery, J. A., Z. Sautner and L. T. Starks, 2016, “Behind the scenes: the corporate governance preferences of institutional investors,” *Journal of Finance* 71:6, 2905-2932

amid financial motivations.⁴ Yet, being universal owners, other researchers claim that these large asset managers can potentially play a pivotal role in reducing climate and other sustainability risks that affect market performance.⁵

The current ESG landscape presents complex dynamics, with research underscoring contrasting trends in the U.S. and Europe.⁶ The controversy surrounding the term “ESG” is significant.⁷ Shifting political sentiments in the U.S. appear to downplay the inclination of The Big Three and institutional investors to exert influence for societal benefit.⁸ Supporters of the “anti-woke” movement perceive ESG as a subjective preference,⁹ contending that pension funds and institutional investors should exclude ESG criteria from their investments.¹⁰ In August 2022, BlackRock CEO, Larry Fink, received a letter from Republican attorney generals, accusing the asset manager of prioritizing its climate agenda over pension beneficiaries’ interests.¹¹ Florida withdrew its assets from BlackRock in protest to Fink’s sustainability statements,¹² and many U.S. states have introduced anti-ESG legislative proposals.¹³ In December 2023, Tennessee sued BlackRock, alleging violations of consumer protection laws through the misuse of ESG factors in its investment strategy.¹⁴ Skepticism about ESG is evident even among financial industry leaders. In 2022, Stuart Kirk, HSBC’s global head of responsible investing, dismissed concerns about climate risk, stating that such risks are too distant for banks to consider and carry minimal financial risk.¹⁵ Kirk’s perspective, shared by many, is that political and financial leaders may overstate the threats posed by climate change and other sustainability risks, viewing ESG primarily as an expression of ideology.¹⁶

“
The current ESG landscape presents complex dynamics, with research underscoring contrasting trends in the U.S. and Europe.
 ”

In Europe, in contrast, a prevailing belief underscores the indispensability of ESG investing and active ESG ownership for fostering a sustainable economy.¹⁷ The core idea is that the financial services sector must channel capital into sustainable investments to ensure enduring economic growth.¹⁸ The pivotal question is not whether ESG should be pursued, but how regulations can be leveraged to amplify sustainable investment activities and engagement by institutional investors, thereby contributing to a more sustainable economic landscape. This distinct European perspective appears to result in a greater commitment to ESG goals among European institutional investors, as evidenced by recent studies and in contrast to their U.S. counterparts.¹⁹

In this article, we delve into the divergent regulatory, political, and societal trends in Europe and the U.S. regarding the rights and duties of institutional investors concerning ESG. Two primary avenues for investors influencing decision making within a company are commonly identified: voice, and exit and selection.²⁰ Shareholders can either directly encourage

⁴ Including, for instance, Bebchuk, L. A. and S. Hirst, 2019, “Index funds and the future of corporate governance: theory, evidence, and policy,” *Columbia Law Review* 119, 2029-2146; Bebchuk, L.A. and S. Hirst, 2022, “Big Three power, and why it matters,” *Boston University Law Review* 102, 1547-1600; Goshen, Z. and A. Hamdani, 2023, “Will systematic stewardship save the planet?” *European Corporate Governance Institute – law working paper no. 739/2023*.

⁵ Including, for instance, Azar, J., M. Duro, I. Kadach and G. Ormazabal, 2021, “The Big Three and corporate carbon emissions around the world,” *Journal of Financial Economics* 142, 674-696.

⁶ ShareAction, 2023, *Voting Matters 2023*

⁷ For a discussion of the history and use of the term ‘ESG’, see Pollman, E., 2022, “The making and meaning of ESG,” *European Corporate Governance Institute – law working paper no. 659/2022*.

⁸ Bebchuk, L. A. and S. Hirst, 2022, “Big Three power, and why it matters,” *Boston University Law Review*, Volume 102, 1547-1600

⁹ Pollman, E., 2022, “The making and meaning of ESG,” *European Corporate Governance Institute – law working paper no. 659/2022*

¹⁰ See, for example Lipton, M., 2022, “ESG, stakeholder governance, and the duty of the corporation,” *Harvard Law School Forum on Corporate Governance* blog dated September 18.

¹¹ See <http://tinyurl.com/mtvym49>.

¹² Master, B., 2023, “BlackRock steps up spending on U.S. lobbying in face of anti-ESG attacks,” *Financial Times* Jan. 29

¹³ Worland, J., 2023, “Lone star ‘wake up call’: Texas Republicans want to ban ESG in insurance,” *Time*, March 1. The article refers to an analysis of anti-ESG laws by *Capital Monitor*, <http://tinyurl.com/zdkjv245>.

¹⁴ Schmitt, W., 2023, “BlackRock sued by Tennessee over ESG strategies,” *Financial Times*, December, 18

¹⁵ See <http://tinyurl.com/42cu6mj8> (around minute 5:05).

¹⁶ Edgecliffe-Johnson, A., 2022, “The war on ‘woke capitalism’,” *Financial Times*, May 27, Pollman, E., 2022, “The making and meaning of ESG,” *European Corporate Governance Institute – law working paper no. 659/2022*

¹⁷ European Commission, 2021, “Strategy for financing the transition to a sustainable economy,” July 6

¹⁸ *Idem*.

¹⁹ Including, for instance, ShareAction, 2023, “*Voting Matters 2023*,”; Lafarre, A. J. F., 2024, “Do institutional investors vote responsibly? Global evidence,” *TILEC discussion paper no. DP2022-001*.

²⁰ Hirschman, A. O., 1970, *Exit, voice and loyalty*, Harvard University Press

corporate management to instigate change or abstain from including the company in their investment portfolio altogether, or opt to exit the company, thereby indirectly impacting corporate management conduct.

2. INVESTMENT STRATEGY AND ESG DUTIES

In response to growing concerns regarding deceptive investor practices, adoption of disclosure rules related to investment strategies has gained attention among regulators. The absence of standardized information in sustainable investing creates a breeding ground for misleading practices,²¹ making uniform disclosure obligations a potential solution.²² These obligations may compel institutional investors to enhance transparency, enabling clients and beneficiaries to compare investment opportunities and make well-informed decisions while encouraging investors to align with sustainability preferences. Consequently, institutional investors may find themselves competing not only on conventional financial factors but also on the sustainability spectrum.²³ This shift allows corporate sustainability leaders to distinguish themselves, garnering reputational benefits and attracting funds from sustainability-focused clients. Many researchers, however, question the effectiveness of such disclosure obligations as they do not directly require institutional investors to change their behavior; hence, their disclosures might reflect nothing more than the status quo.²⁴ Others highlight the complexity of sustainable finance information.²⁵ Notably, *The Economist* magazine highlights the challenges that ESG rating agencies face, indicating measurement problems that lead to contradictory scores, often forming the foundation of sustainable investment strategies.²⁶ Notwithstanding these limitations, there remains a regulatory focus on ESG disclosure obligations on both sides of the Atlantic.

2.1 ESG (disclosure) duties in the U.S.

In the U.S., there is a general movement towards more reliable sustainability information. On May 25, 2022, the U.S. Securities and Exchange Commission (SEC) proposed new disclosure requirements for ESG funds. First, there would be three categories of registered ESG funds: (1) “integrated” (funds that consider ESG factors, but those factors are not the primary consideration), (2) “focused” (ESG factors are the primary consideration) and (3) “impact” (funds that pursue ESG impact).²⁷ The proposal also requires ESG-focused funds that claim to consider environmental issues to include GHG (greenhouse gas) emissions data related to their portfolio company investments unless the fund discloses that it does not consider GHG emissions as part of its investment strategy.²⁸ Previously, the SEC had proposed requiring large companies to report on climate-related risks and GHG emissions.²⁹ In another proposal, approved on September 2023, the SEC proposed to modify the scope of the “Names Rule”, which states that if a fund’s name suggests a particular focus, at least 80% of the value of its assets must be invested accordingly – to include funds using ESG-related names.³⁰

Although these SEC proposals seem to indicate that the U.S. is heading towards more ESG duties for institutional investors, this trajectory is not without political debate. Under the Trump administration, the U.S. Department of Labor (DOL) had proposed a change in the law to allow pension funds governed by the Employee Retirement Income Security Act of 1974 (ERISA) to include only “pecuniary factors” in their investment decisions as part of their fiduciary duty.³¹ The final version of this law states that an investment decision must be based solely on monetary factors and to not subordinate the interests of participants and beneficiaries to non-monetary objectives.

²¹ Berg, F., K. Fabisik, and Z. Sautner, 2021, “Is history repeating itself? The (un)predictable past of ESG ratings,” *European Corporate Governance Institute – finance working paper no. 708/2020*

²² Paccès, A., 2021, “Will the EU Taxonomy Regulation foster sustainable corporate governance?” *Sustainability* 13:21, 12316

²³ *Idem*.

²⁴ Including, for instance, Bruner, C., 2022, “Corporate governance reform and the sustainability imperative,” *Yale Law Journal* 131:4.

²⁵ Ahlström, H. and B. Sjöfjell, 2022, “Complexity and uncertainty in sustainable finance: an analysis of the EU taxonomy,” in Cadman, T. and T. Sarker (eds.), *De Gruyter handbook of sustainable development and finance*, De Gruyter

²⁶ *The Economist*, 2022, “ESG investing. A broken idea,” July 21, <http://tinyurl.com/yrvzrk4x>. Following Cools, S., 2023, “Climate proposals: ESG shareholder activism sidestepping board authority,” in Kuntz, T., (ed.), forthcoming, *Research handbook on environment, social, and corporate governance*, Edward Elgar Publishing.

²⁷ Funds that do not take ESG factors into account are not rated.

²⁸ See <http://tinyurl.com/26d7ny4y>

²⁹ See <http://tinyurl.com/4y6dws6w>

³⁰ See <http://tinyurl.com/bdfbmc86>. For a discussion of the Names Rule, see: Fisch, J. E. and A. Z. Robertson, 2023, “What’s in a name? ESG mutual funds and the SEC’s Names Rule,” *European Corporate Governance Institute – law working paper no. 697/2023*.

³¹ DOL, 72846 Federal Register 85(220), November 13, 2020, available at <http://tinyurl.com/4f4w5at5>. For the 2020 law, see <http://tinyurl.com/3uydt4rh>.

DOL does recognize that ESG factors may be compatible with a purely financial analysis of an investment decision. Non-monetary objectives can serve as a “tie-breaker” if investment options are financially indistinguishable, but this requires documentation of why the monetary factors were insufficient to make the decision, including a comparison of investment options and how the non-monetary objectives are consistent with the financial interests of participants and beneficiaries.

In November 2022, however, DOL passed new legislation under the Biden administration (“DOL’s ESG Rule”).³² This ESG Rule removed the term “pecuniary factors” and emphasizes that investment decisions focus on the relevant “risk-return factors”, and that ESG factors may be included here. DOL states that the new law seeks to eliminate “the chilling effect created by the prior administration on considering environmental, social and governance factors in investments.”³³ In essence, DOL’s ESG Rule from 2022 does not differ that much from the 2020 one, and does not really encourage the consideration of ESG factors.³⁴ However, it does remove the ambiguity as to whether the inclusion of ESG factors is permissible and the administrative costs that accompanied it under the Trump administration’s legislation. Particularly, DOL’s ESG Rule confirms that when selecting investments, pension funds must focus on relevant risk-return factors and not subordinate the interests of participants and beneficiaries to objectives unrelated to benefits within a pension plan. Republicans (and two Democrats) stopped this law in early March 2023 on the grounds that it would be part of woke capitalism, after which President Biden used his veto power – for the first time – on March 20, 2023 against this Congressional resolution.³⁵ Adding to the ambiguity surrounding the status of the DOL’s ESG Rule is the filing of several lawsuits against DOL aiming to prevent its enforcement.³⁶

2.2 ESG (disclosure) duties in Europe

In recent years, there has been a significant surge in emphasis on sustainable finance and corporate governance within the European Union. Europe is actively pursuing this goal through its 2018 Sustainable Finance Action Plan and its renewed strategy for financing the transition to a sustainable economy,³⁷ primarily relying on transparency obligations outlined in regulatory initiatives such as the Sustainable Finance Disclosure Regulation (SFDR),³⁸ Corporate Sustainability Reporting Directive (CSRD),³⁹ and the Taxonomy Regulation,⁴⁰ among others.⁴¹ The SFDR plays a pivotal role in clarifying institutional investors’ responsibilities regarding sustainability. It mandates financial market participants to furnish detailed information about sustainability risks, the sustainable attributes of financial products, and their adverse impacts on sustainability factors. One notable feature is the SFDR’s classification of ESG funds, ranging from Article 6 (no sustainability objective) to Article 8 (fostering sustainability characteristics, light-green), and Article 9 (with a sustainability objective, dark-green). Complementing the SFDR are technical standards (RTS) presented as delegated regulations, offering additional insights into the content and methodology of disclosure requirements.⁴²

Notably, the latest updates to the RTS, focusing on sustainable investments in the fossil gas and nuclear sectors, came into effect on February 21, 2023.⁴³ Moreover, as part of Europe’s sustainable financial strategy, revisions to the MiFID II Delegated Regulation⁴⁴ necessitate investment firms to incorporate their clients’ sustainability preferences into the advisory process. These adjustments mandate investment firms to ensure that transactions align with their clients’ investment objectives, encompassing both risk tolerance and sustainability preferences.

³² For this 2022 law, see <http://tinyurl.com/yc8yy8p3>.

³³ See Dyer, E., M. Albano, C. Gottlieb, 2022, “New DOL guidance on ESG and proxy voting,” Harvard Law School Forum on Corporate Governance blog, December 22.

³⁴ See Malone, L., E. Rozow, and G. M. Gerstein, 2023, “Biden’s first veto: understanding the implications of the DOL’s ESG rule,” Harvard Law School Forum on Corporate Governance blog, April 6.

³⁵ Gardner, A., 2023, “Biden vetoes bill for first time to block anti-ESG measure,” Bloomberg, March 20. See also, Fedor, L. and J. Politi, 2023, “Joe Biden expected to issue first presidential veto in anti-ESG vote,” Financial Times, March 1.

³⁶ See Malone, L., E. Rozow, and G. M. Gerstein, 2023, “Biden’s first veto: understanding the implications of the DOL’s ESG rule,” Harvard Law School Forum on Corporate Governance blog, April 6.

³⁷ European Commission, 2021, “Strategy for financing the transition to a sustainable economy,” July 6

³⁸ Regulation (EU) 2019/2088

³⁹ Directive 2022/2464/EU

⁴⁰ Regulation (EU) 2020/852

⁴¹ There is also a proposed regulation for a standard for European green bonds dated July 6, 2021 (also called “European green bonds” or “EuGBs”) that was adopted by the Council in October 2023.

⁴² Delegated Regulation (EU) 2022/1288

⁴³ Delegated Regulation EU 2023/363

⁴⁴ Delegated Regulation (EU) 2021/1253 of 21 April 2021 amending Delegated Regulation (EU) 2017/565 as regards the integration of sustainability factors, risks and preferences into certain organizational requirements and operating conditions for investment firms.

The overview presented above highlights Europe's commitment to transparency obligations, including the advisory process, and the uniformity of sustainability disclosures within capital markets. Despite the complexity and ongoing changes in the European framework,⁴⁵ these obligations are designed to contribute significantly towards the actual sustainability of ESG investments. Clients and beneficiaries of institutional investors are empowered to compare investment options, facilitating well-informed investment decisions. Ideally, this shift will prompt institutional investors to compete not only on traditional financial returns but also on the sustainability of their investments.⁴⁶

Can the direction set by the European legislature yield the intended results? Some scholars have expressed skepticism. Bruner (2022), for instance, contends that while transparency is often viewed as a crucial precursor to meaningful reform, it is frequently treated as a substitute for it.⁴⁷ Additionally, the question remains whether less sustainable companies will genuinely face a higher cost of capital.⁴⁸ In these cases, active ownership remains the preferred option. However, research shows that these ESG disclosure duties have some positive effects. For instance, Dai et al. (2023) study the effects of the SFDR and find that it has triggered a significant decarbonization of the investment portfolios within E.U. funds professing a commitment to sustainability criteria.⁴⁹ According to the authors, these reduced emissions levels can be attributed to both alterations in funds' investment strategies and shifts in firm-level emissions. It seems that with disclosure duties like the SFDR institutional investors have the ability to truly signal that they are investing sustainably. Ideally, greenwashing practices become more challenging, fostering a genuine emphasis on sustainability.

The unfolding European initiatives present contrasting trajectories compared to trends in the U.S., especially concerning the ongoing discourse on the compatibility of ESG investing and fiduciary duties under ERISA in the latter. However, even in Europe, there is an ongoing debate regarding ESG obligations of financial services organizations. Notably, the provisional agreement on the Corporate Sustainability Due Diligence Directive (CSDDD),⁵⁰ dated December 14, 2023,⁵¹ underscores that while the financial services sector is encompassed in the legislative initiative, its application will be limited. Specifically, financial entities will only be required to implement the CSDDD for a limited part of their supply chains.

3. ACTIVE OWNERSHIP

Within the exit-voice dichotomy, voice is widely acknowledged to be the more powerful tool.⁵² Shareholder engagement, often viewed as a form of shareholder activism, involves shareholders proactively initiating meaningful dialogues, frequently conducted discreetly behind the scenes.⁵³ Additionally, investors can exercise their formal voice rights, such as voting and shareholder proposal rights.⁵⁴ In this section on active ownership, the analysis focuses on shareholder sustainability voting, as voting serves as a crucial escalation strategy for institutional investors to exert influence on corporate management.⁵⁵

3.1 Active ownership in the U.S.

While shareholder activism in the U.S. has historically been associated with small individual shareholders, known as "corporate gadflies", and more aggressive hedge funds, who dominate the agendas of large corporations with their

⁴⁵ For instance, Partiti, E., 2023, "Addressing the flaws of the Sustainable Finance Disclosure Regulation: moving from disclosures to labelling and sustainability due diligence," forthcoming in European Business Organisation Law Review.

⁴⁶ Paccas, A., 2021, "Will the EU Taxonomy Regulation foster sustainable corporate governance?" Sustainability 13:21, 12316

⁴⁷ Bruner, C., 2022, "Corporate governance reform and the sustainability imperative," Yale Law Journal 131:4

⁴⁸ Anabtawi, I. and L. Stout, 2008, "Fiduciary duties for activist shareholders," Stanford Law Review 60:5, 1255-1308

⁴⁹ Dai, J., G. Ormazabal, F. Penalva, and R. A. Raney, 2023, "Imposing sustainability disclosure on investors: does it lead to portfolio decarbonization?" European Corporate Governance Institute – finance working paper 945/2023

⁵⁰ Proposal for a Directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive (EU) 2019/1937, February 23, 2022

⁵¹ See <http://tinyurl.com/227jn3f9>

⁵² Broccardo, E., O. Hart, and L. Zingales, 2020, "Exit vs. voice," ECGI-Finance working paper no. 694/2020

⁵³ McCahery, J. A., Z. Sautner, and L. T. Starks, 2016, "Behind the scenes: The corporate governance preferences of institutional investors," Journal of Finance 71:6, 2905-2932

⁵⁴ Grewal, J., G. Serafeim, and A. Yoon, 2016, "Shareholder activism on sustainability issues," Harvard Business School Working Paper, No. 17-003; Lee, M.-D. and M. Lounsbury, 2011, "Domesticating radical rant and rage: an exploration of the consequences of environmental shareholder resolutions on corporate environmental performance," Business & Society 50:1, 155-188

⁵⁵ Lafarre, A. J. F., 2024, "Do institutional investors vote responsibly? Global evidence," TILEC discussion paper no. DP2022-001

shareholder proposals, there has been a noticeable shift in recent years. Institutional investors, who nowadays own the majority of shares in companies worldwide, are no longer remaining silent and have instead started to support smaller activists and combine their powers in collaborative engagements using shareholder proposals.

A prime example of this shift is the unprecedented success of a newcomer activist group called Engine No. 1 in its proxy fight with ExxonMobil. Launched in December 2020 as an “impact hedge fund”,⁵⁶ Engine No. 1 nominated four independent director candidates to the board of directors of ExxonMobil at the 2021 AGM. Despite owning only 0.02% of Exxon Mobil’s stock, the fund was able to oust and replace three directors with the help of institutional investors. Engine No. 1’s example also highlights another shift, namely the shift from proposals being mostly focused on governance issues – such as plurality voting rules, staggered boards, protection mechanisms, and access to the company’s proxy – to shareholder proposals on sustainability topics, particularly climate change. In recent years, we have witnessed an increase in number of shareholder proposals submitted, with the highest level of submissions since 2016 in 2023.⁵⁷

Regulations set forth by the SEC empower boards to exclude certain proposals from a company’s proxy materials. These exclusions typically pertain to matters deemed inappropriate under state law or those concerning the company’s routine business operations, as outlined in section 14a-8 of the Securities Exchange Act. The focal point of these no-action reliefs commonly revolves around the ordinary business operations ex Rule 14a-8(i)(7). The SEC employs a two-fold approach to evaluate the exclusion eligibility of a proposal under this exception. Firstly, a matter can be excludable for relating to ordinary business if it is fundamental to management’s ability to run a company on a day-to-day basis that the matter could not, as a practical matter, be subject to direct shareholder oversight. In the Staff Legal Bulletin from October 2019,⁵⁸ the SEC, however, indicated that a company will not be permitted to exclude a proposal based on this ground that transcends the day-to-day business operations because it raises “a policy issue so significant” that it would be appropriate for a shareholder vote.

Climate-related resolutions are often categorized as significant enough to warrant the latter exception of a significant policy issue. This trend has been accentuated, particularly for climate proposals, since the end of 2021: the SEC announced its decision to no longer necessitate shareholders to demonstrate the issue’s significance to the “specific” company.

Secondly, the SEC considers shareholder proposals that “excessively micro-manage the company” as related to ordinary business operations. In the same Bulletin, the SEC explained that a shareholder proposal may be considered micromanaging if it is too prescriptive, limiting the discretionary powers of the board of directors. As a result, the SEC excludes proposals that prescribe emission reduction targets in an overly detailed manner, such as stipulating specific methods for establishing or achieving these targets. The SEC illustrates the dichotomy in its approach by citing two sample shareholder proposals related to environmental, social, and governance (ESG) matters:

- **Proposal 1:** a proposal on annual reporting about “short-, medium- and long-term greenhouse gas targets aligned with the greenhouse gas reduction goals established by the Paris Climate Agreement to keep the increase in global average temperature to well below 2 degrees Celsius and to pursue efforts to limit the increase to 1.5 degrees Celsius.”⁵⁹
- **Proposal 2:** a proposal requesting a report “describing if, and how, [a company] plans to reduce its total contribution to climate change and align its operations and investments with the Paris [Climate] Agreement’s goal of maintaining global temperatures well below 2 degrees Celsius.”⁶⁰

The first proposal, characterized by its excessive level of prescription, can be excluded. Conversely, the second proposal, characterized by its more general nature, would not be subject to exclusion.

While the SEC’s more lenient stance on climate-related proposals may have led to an increase in ESG proposals, there is a concurrent tightening of thresholds. Under Rule 14a-8, shareholders were previously eligible to request the inclusion of their proposals in proxy materials if they held a minimum of

⁵⁶ Christie, A., 2021, “The agency costs of sustainable capitalism,” UC Davis Law Review 55, 875-954

⁵⁷ See, Mueller, R. O., E. A. Ising, and T. J. Kim, 2023, “Shareholder proposal developments during the 2023 proxy season,” Harvard Law School Forum on Corporate Governance blog, August 3.

⁵⁸ Staff Legal Bulletin No. 14K (CF) (SLB No. 14K)

⁵⁹ Devon Energy Corp. (March 4, 2019)

⁶⁰ Anadarko Petroleum Corp. (March 4, 2019)

U.S.\$2,000 market value or 1% of the company's voting shares for at least one year preceding the submission. However, as of January 1, 2022, the threshold underwent a significant shift, becoming more contingent on the duration of shareholding. Shareholders now need U.S.\$2,000 worth of the company's shares if held for a minimum of three years, U.S.\$15,000 worth for a holding of at least two years, or U.S.\$25,000 worth for a holding duration of at least one year. This adjustment reflects a more stringent criterion for shareholders seeking to include their proposals in the company's proxy materials.

In addition, the amendments that – transitionally – entered into force for shareholders on January 1, 2023 impose a one-proposal limit on “each person” for shareholder meetings, meaning a proponent can submit only one proposal, regardless of their capacity as a shareholder or a representative. Regarding resubmissions of shareholder proposals, the amendments raise the thresholds for excluding proposals addressing the same subject within the past five years to 5%, 15%, and 25% for votes received on matters previously voted on once, twice, or three or more times, respectively. These amendments imposed stricter rules for shareholders to submit shareholder proposals. But stricter rules will likely also apply on companies seeking to exclude shareholder proposals: on July 13, 2022, the SEC proposed further amendments to the Shareholder Proposal Rule 14a-8, but this time stricter requirements are put on companies seeking no-action relief.⁶¹ Particularly, the suggested amendments aim to heighten the criteria for three key grounds of exclusion, making reliance on substantial implementation, duplication, and resubmission grounds for exclusion more challenging.

3.2 Active ownership in Europe

The European Commission (E.C.) addressed corporate governance shortcomings exposed by the global financial crisis, particularly the inadequate engagement of institutional investors. The 2012 Action Plan⁶² led to the E.C.'s announcement of a package to enhance shareholder engagement and corporate governance reporting, culminating in the adoption of the revised shareholder rights directive (SRD II) in 2017.⁶³

The Preamble of SRD II emphasizes shareholder engagement as a fundamental aspect of corporate governance, asserting that increased shareholder involvement can enhance both the financial and non-financial performance of companies, including factors related to environmental, social, and governance (ESG). The Directive operates under the corporate governance principle that shareholders play a crucial role in holding management accountable for their actions.⁶⁴ Articles 3g-3i of SRD II outline institutional investors' duties, including the disclosure of an engagement policy, monitoring of investments on crucial matters, dialogue with investee companies, exercising voting rights, cooperating with shareholders and stakeholders, and addressing conflicts of interest. The comply-or-explain principle applies to these obligations, such as disclosing the implementation of the policy and characteristics of arrangements with asset managers. Article 3h focuses on the alignment of investment strategy with long-term liabilities, and Article 3i requires asset managers to disclose how their strategy aligns with institutional investors' interests, promoting informed selection and alignment of long-term interests. Hence, following SRD II, but also the many stewardship codes that are adopted by European member states and other countries,⁶⁵ institutional investors are increasingly expected to showcase their proactive use of shareholder rights for sustainability purposes.

However, despite these initiatives, a significant gap remains between the SRD II framework and the national corporate laws of the European member states. The limitations imposed by member states' laws, grounded in the autonomy of boards, hinder the framework's ability to fully meet Europe's expectations. In traditional corporate governance discussions, two legal approaches are commonly discussed: regulatory strategies that limit the actions of company agents and governance strategies that empower shareholders (the principals).⁶⁶ While it is often believed that European member states typically adopt a governance strategy more often than the U.S., when it comes to sustainability engagement, it appears that shareholder rights in Europe are lagging behind.

⁶¹ See: <http://tinyurl.com/mswerp2k>

⁶² Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Action Plan: European company law and corporate governance – a modern legal framework for more engaged shareholders and sustainable companies (December 12, 2012)

⁶³ Directive (EU) 2017/828

⁶⁴ European Commission, 2011, “The EU Corporate Governance Framework,” European Commission Green Paper COM(2011) 164 final, October 27

⁶⁵ Katalouzou, D. and D. W. Puchniak, 2022, *Global shareholder stewardship*, Cambridge University Press

⁶⁶ Kraakman, R., J. Armour, P. Davies, L. Enriques, H. Hansmann, G. Hertig, K. Hopt, H. Kanda, M. Pargendler, W-G. Ringe, and E. Rock, 2017, *The anatomy of corporate law: a comparative and functional approach*, third edition, Oxford University Press

The involvement of investors in sustainability matters is hindered by the distribution of substantive powers outlined in national corporate law statutes in Europe.⁶⁷ In many instances, corporate law systems in Europe categorize topics falling under the ESG umbrella as strategic matters within the purview of the board of directors, not the shareholder meeting. The scarcity of shareholder proposals in Europe is linked to regulatory differences in shareholder engagement and ownership disclosure, as well as distinctive stock ownership structures. Additionally, there is a perceived lower demand for activism on issues that have traditionally been more prominent in the U.S.⁶⁸ In the Netherlands, for instance, shareholder proposals face restrictions due to “oligarchic clauses” commonly found in the articles of association of Dutch listed companies.⁶⁹ Such clauses, for instance, necessitate shareholder resolutions to obtain approval from the managing or supervisory board, limiting the autonomy of shareholders. In addition, Dutch case law has solidified a doctrine emphasizing strong board autonomy,⁷⁰ making it practically impossible for shareholders to introduce binding and non-binding proposals related to the board’s competence at shareholder meetings without the board’s permission.

An important example of ESG proposal restrictions in France can be found at the 2022 AGM of the oil major TotalEnergies. In 2022, a consortium of institutional investors proposed a climate shareholder resolution to be included in the agenda of TotalEnergies’ shareholder meeting. Following article L 225-105(2) of the “French Commercial Code” (FCC), one or more

shareholders that represent at least 5% of the capital have the right to add a shareholder resolution to the shareholder meeting’s agenda.⁷¹ TotalEnergies’ corporate board, however, refused to put it to a vote, arguing that the shareholder meeting is not the competent corporate body to decide on such a strategy matter.⁷² Some members of the consortium voted against the re-election of TotalEnergies’ board members in response.⁷³ The institutional investors formed again a consortium in 2023 to file another climate resolution at TotalEnergies’ shareholder meeting. To ensure that the climate resolution will not be refused from TotalEnergies’ 2023 AGM’s agenda, the investors decided to make the resolution a consultative (non-binding) one.⁷⁴

The inclusion of climate or broader sustainability-related shareholder proposals on the agenda emerges as a crucial element in steering the financial transition towards more sustainable business practices. While establishing direct causality remains challenging, research indicates that such proposals can exert a positive influence on corporate sustainability performance. Notably, Flammer et al. (2021) revealed that climate proposals contribute to companies’ increased voluntary disclosures of climate risks.⁷⁵ Grewal et al. (2016) established a connection between shareholder proposals and ESG performance.⁷⁶ Additional insights from Bauer et al. (2022) sheds light on the dynamics of successful shareholder proposals. Their research underscores that success is not solely measured by actual votes but also by the withdrawal of proposals following fruitful discussions with the board.⁷⁷

⁶⁷ Cools, S., 2023, “Climate proposals: ESG shareholder activism sidestepping board authority,” in Kuntz, T., (ed.), forthcoming, Research handbook on environment, social, and corporate governance, Edward Elgar Publishing

⁶⁸ Idem.

⁶⁹ Kemp, B., 2020, “Limiting shareholder power in Dutch listed companies,” Oxford Business Law blog of May 21

⁷⁰ Including HR 4 April 2014, NJ 2014, 286 (Cancun); OK May 29, 2017, JOR 2017/261 (AkzoNobel), HR 20 April 2018, ECLI:NL:HR:2018:652 (Boskalis / Fugro).

⁷¹ If the capital is €750,000 or lower. Note that the threshold progressively declines with the size of the company’s capital ex article R225-71. For instance, the threshold is 1 percent for the portion of capital between €7,500,000 and €15,000,000, and 0.50 percent for a larger share capital.

⁷² In 2022, the Haut Comité Juridique de la Place Financière de Paris – a committee created to address legal uncertainties surrounding climate resolutions – confirmed that the climate strategy falls within the board’s statutory competence to “set out the orientation” of the company. Haut Comité Juridique, Rapport sur les résolutions climatiques ‘Say on climate’ 13 (December 15, 2022), available at <http://tinyurl.com/2tj4mwx3>. As a result, it seems that resolutions that mandate the board to achieve certain emissions targets or hold a shareholders’ vote on climate issues may infringe on the board’s powers and thus may be excluded from the agenda of a shareholders’ meeting. Following Cools (2023).

⁷³ For instance, MN stated that: “[w]e cannot approve the re-election of the board members, since we hold the board members responsible for denying a shareholder proposal being added to the ballot.” And Kempen Capital Management announced that: “[w]e would like make use of this opportunity, to express our dissatisfaction with TotalEnergies’ management reluctance to place a resolution that we have co-filed with a group of other share- / stakeholders asking the company to set short, medium and long-term targets to limit climate change in line with the Paris Climate Agreement.” Voting rationales retrieved from the Insightia database on 1 February 2024.

⁷⁴ See for more information about this resolution: <http://tinyurl.com/593fk2zm>.

⁷⁵ Flammer, C., M. W. Toffel, and K. Viswanathan, 2021, “Shareholder activism and firms’ voluntary disclosure of climate change risks,” Strategic Management Journal 42:10, 1850-1879

⁷⁶ Grewal, J., G. Serafeim, and A. Yoon, 2016, “Shareholder activism on sustainability issues,” Harvard Business School Working Paper, No. 17-003. See also, Lee, M-D. and M. Lounsbury, 2011, “Domesticating radical rant and rage: an exploration of the consequences of environmental shareholder resolutions on corporate environmental performance,” Business & Society 50:1, 155-188; Bauer, R., J. Derwall, and C. Tissen, 2022, “Corporate directors learn from environmental shareholder engagements,” SSRN

⁷⁷ Bauer, R., J. Derwall, and C. Tissen, 2022, “Corporate directors learn from environmental shareholder engagements,” SSRN; Bauer, R., F. Moers, and M. Viehs, 2015, “Who withdraws shareholder proposals and does it matter? An analysis of sponsor identity and pay practices,” Corporate Governance: An International Review 23(6), 472-488

Crucially, the absence of the right for shareholders to submit competing climate proposals to shareholder meetings could potentially skew management's understanding of shareholder preferences, particularly in the context of "Say on climate" initiatives. An example is the Shell 2021 AGM, where approximately 89% of shareholders endorsed the management's climate proposal, despite it not aligning with the Paris Agreement. This seemingly high level of support might mislead observers into thinking that the majority of shareholders endorse Shell's climate strategy. However, over 30% also supported the competing Follow This climate proposal, advocating for a stricter and Paris-aligned climate strategy.⁷⁸ This disparity highlights that a significant portion of shareholders had reservations about Shell's climate plans, contrary to what the management proposal suggested. The right for shareholders to present alternative proposals can be pivotal in ensuring a comprehensive and accurate representation of shareholder sentiments on critical issues such as climate strategy.

Moreover, in addition to the constraining doctrines imposed by European member states, the rules governing collaborative actions in Europe further complicate concerted sustainability engagement efforts for institutional investors. This complexity becomes evident, for example, when investors seek to coordinate their votes in support of a climate resolution, potentially triggering the obligation to launch a public offer for all remaining shares. This uncertainty poses challenges for collaborating investors, raising questions about the extent of their cooperative actions.⁷⁹ Addressing these concerns, in 2013, the European Securities and Markets Authority (ESMA) introduced a "white list" delineating activities in which shareholders could collaborate without being automatically presumed to have acted in concert.⁸⁰ Recognizing the evolving landscape of sustainability considerations, ESMA initiated an evaluation of this framework in 2019. The objective is to determine whether the existing guidance might be overly restrictive for institutional investors collaborating, particularly in the context of addressing ESG matters.



⁷⁸ The Follow This resolution can be found here: <http://tinyurl.com/3v47seda>.

⁷⁹ Article 2.1(d) of the Takeover Bids Directive defines "persons acting in concert".

⁸⁰ ESMA, 2013, "Public statement containing information on shareholder cooperation and acting in concert under the Takeover Bids Directive," ESMA/2013/1642, <http://tinyurl.com/3k7kp2kw>

4. CONCLUSION

In conclusion, this research underscores the significant divergence in regulatory, political, and societal trends between Europe and the U.S. concerning the ESG rights and duties of institutional investors. Although the SEC demonstrates an inclination towards heightened ESG duties, this trajectory is not devoid of political debate. Notably, despite the SEC's commitment to ESG transparency, the U.S. grapples with the fundamental question of whether sustainability should be pursued, often reducing ESG discussions to mere ideology. In contrast, Europe has witnessed a significant surge in emphasizing sustainable finance and corporate governance. The European focus centers on transparency obligations outlined in various regulatory initiatives, including the SFDR. This European approach starkly contrasts with ongoing debates in the U.S., particularly regarding the compatibility of ESG investing and fiduciary duties under ERISA.

In terms of active ownership and shareholder voting, the U.S. has seen institutional investors actively supporting smaller activists and engaging in collaborative efforts using shareholder proposals, perhaps partly driven by the SEC's more lenient stance on climate-related proposals. However, in Europe, despite the strong emphasis on sustainable finance, the national frameworks of member states do not align with European goals, necessitating a reevaluation. To bridge these gaps and cultivate a more harmonized and effective approach to sustainable investment, we advocate for aligning European aspirations for capital allocation with an increased emphasis on sustainability voice in member states, potentially through the forthcoming proposal for the next Shareholder Rights Directive (SRD III). This Directive could specifically aim at harmonizing European member states' rules with the European Green Deal framework, particularly in terms of institutional investor ESG duties. The introduction of a Say-on-Climate mechanism and a concerted effort to amplify shareholder voice within member states can substantially contribute towards aligning European goals for capital allocation with sustainable investments.

HOW BANKS RESPOND TO CLIMATE TRANSITION RISK

BRUNELLA BRUNO | Tenured Researcher, Finance Department and Baffi, Bocconi University

ABSTRACT

We investigate whether and how banks in the global syndicated loan market adjusted the pricing and supply of credit to account for higher climate transition risk. We provide a comprehensive measure of exposure to climate transition risk, considering three important risk drivers: the borrower's carbon emissions, a policy shock represented by the 2015 Paris Agreement, and climate resilience and policy stringency of the country in which borrowers are located. The evidence is mixed and points to non-linear relations between lending variables and CO₂ emissions. Policy events such as the Paris Agreement and government environmental awareness are significant climate risk drivers that, when combined, may amplify banks' perception of climate transition risk.

1. INTRODUCTION

Coping with climate risks, whether they are physical or transition-related, has become a priority for various stakeholders in the financial services sector. Banks, particularly, play a unique role, because the success of the transition toward a greener economy depends on how effectively they can channel credit towards low-emission borrowers and industries.

Climate change impacts bank balance sheets through macro- and microeconomic transmission channels stemming from two distinct types of climate risk drivers. First, banks may incur economic costs and financial losses due to the escalating severity and frequency of physical climate risk drivers. Second, they may be affected by how shifts in government policies, technological advancements, and changes in investor and consumer sentiment steer the economies' efforts in curtailing carbon emissions. In both scenarios, increased climate risk can manifest directly through banks' exposures to borrowers and countries facing climate-related shocks, or indirectly through the repercussions of climate change on the broader economy and the feedback effects within the financial system. The impacts of climate risk drivers on banks can be observed through "traditional" risk categories, as they become evident through amplified default risks in loan portfolios or decreased values of assets.

A mechanism by which climate change affects bank balance sheets is through the lending channel. To explain this mechanism, increased physical risk may directly impact businesses and households. Extreme weather events can damage properties and other physical assets, as well as impair agricultural productivity and human labor. Consequently, banks more exposed to these households and businesses may suffer from increased default rates and collateral deterioration. Regarding transition risk, the adoption of mitigation policies and changes in sentiment toward climate change may impact polluting companies' businesses through asset stranding, property deterioration, and higher capital expenditure due to transitioning. Once again, banks exposed to industries and businesses more involved in the transition process may experience increased credit losses.

If banks hold climate sentiments, meaning they form expectations about the impact of climate change on their exposures, they could in principle adjust their investment decisions by reallocating resources across borrowers and industries, thereby influencing the outcome of the transition.

In practice, however, there are several factors that make banks' reaction to climate risk hard to predict. First, it is unclear whether models commonly used by banks to measure credit risk are actually able to capture tail-events related to

the repercussions of environmental issues on bank balance sheets. This is partly due to the challenge of quantifying climate change risk, especially when referring to the risks of transitioning to a lower-carbon economy.

Second, perceptions of climate change risk may be intertwined with the credibility of climate policy implementation. For example, delays in enforcing climate policies and policy inconsistencies may affect how climate-related financial risks are perceived. This, in turn, could influence banks' propensity to invest in carbon-intensive firms.

Third, bank investors and stakeholders may prioritize maximizing returns over environmental concerns, as the recent expansion of anti-environmental, social and governance (ESG) laws in certain U.S. states suggests [Donefer (2023)]. As a consequence, instead of promoting it, the banking system may actually hinder the green transition by impeding the financing of innovation in industries most exposed to green technology externalities.

All these explanations underline the fact that the evidence on whether banks incorporate climate risk in their lending decisions is far less clear than the evidence regarding the pricing of climate risk in bond and stock markets [see, for example, Bolton and Kacperczyk (2021)].

2. CLIMATE TRANSITION RISK AND BANK LENDING

2.1 Research questions and the problem of measuring climate transition risk

Bruno and Lombini (2023) contribute to the debate on the role played by banks in coping with climate-related issues by investigating whether and how they adjust the price and amount of credit in reaction to amplified climate change risk. Do banks apply higher interest rates on riskier borrowers and industries? Do they curtail lending to these borrowers and industries?

To address these questions, we focus on climate transition risks, which pertain to the challenges associated with the adjustment process towards a low-carbon economy. This is important because most existing research on climate risk in banking is either qualitative in nature or interested in the effects of physical risks.

The scarcity of empirical evidence on climate transition risks mainly deals with the challenge of measuring banks' and borrowers' exposure to climate transition. The difficulty arises because of the multiple risk drivers influencing the

“
Policy shocks, such as the Paris Agreement and government commitments to environmental issues, are important climate risk drivers that, when combined, amplify banks' perception of transition risk.
 ”

intensity of bank balance sheet exposure to climate risks [BIS (2021)]. First, not only firms but also economic sectors may have different sensitivities towards the transition to a low-carbon economy. Second, climate transition risks can get ignited by specific macro-events (such as changes in government policies and technological improvements) that can either mitigate or exacerbate a single firm's and industry's exposure to the risk of transition. Third, the same macro-shock may affect differently companies and industries based on the geographic locations of either banks or their borrowers. For example, a country's specific commitment to climate-related issues can make the same climate goals potentially more compelling, and related actions more incisive, than in other countries.

To account for multiple risk drivers and interactions that are inherent to climate transition risks, we provide a three-pronged, comprehensive measure of exposure to climate transition risk that encompasses (1) carbon emissions at the borrower levels, (2) a macro-policy shock, and (3) an indicator of a country's commitment to engaging with climate change issues.

The underlying idea of using carbon emissions as a first proxy of borrower exposure to climate transition risk is that more polluting firms are more likely to be targeted by climate regulation, which may entail costs and losses for banks as a result of the mechanism illustrated in the previous section.

The macro-policy shock we exploit in the empirical analysis is the ratification of the Paris Agreement at the closing of the 21st Conference of the Parties (COP21) on December 12th, 2015, an event commonly regarded as a major spark of climate transition risk. The Agreement, which brought together 194 Parties, set out a global framework to avoid dangerous climate change, in the ambitious attempt to reach

climate-neutrality before the end of the century. The best-known resolution of the Agreement is the one related to mitigation policies, meaning actions concerning the reduction of greenhouse gas (GHG) emissions to limit global warming. To achieve this goal, countries have agreed to review their own commitments every five years, as well as to provide financing to developing countries to mitigate climate change and strengthen resilience to adapt to climate impact. With its entry into force on November 4th, 2016, the Paris Agreement became the first-ever universal and legally binding climate change agreement on a global basis.

2.2 Sample and data

We collect bank-firm data from the global loan syndication market, along with firm-level CO₂ emissions data, to measure bank exposures to large corporations across various industries and countries showing broad cross-sectional heterogeneity between green and brown firms.

We rely on multiple sources of data. We retrieve data on syndicated loans from Thomson Reuters DealScan. The unit of observation is the loan (or facility), which is usually grouped into deals or packages. Loan data include details on the lender (name and loan share), the loan (maturity, amount, cost, origination date, presence of collateral, and covenants), and the borrower (name and location). We use this data to construct our lending variables, namely the cost (basis points) and amount (as logarithm of total amount and as a share of total loans) of syndicated loans granted by a given bank to a specific borrower in a year.

We then employ a few direct and indirect indicators of firms' and countries' vulnerability to transition risk. We measure firm-level pollution through the total annual amount of CO₂ emissions (in thousands of tons), as retrieved from Thomson Reuters Eikon, which provides data on total CO₂ emissions (in tons) along with Scope1, Scope2, and Scope3 CO₂ emissions.

In order to capture information on government environmental awareness, we resort to Germanwatch's Climate Change Performance Index (CCPI), which tracks the countries' efforts to combat climate change.¹ This indicator is considered a long-standing and reliable tool for identifying leaders and laggards in climate protection [Delis et al. (2023)]. The CCPI is published annually and gathers several dimensions that are relevant for a country's engagement with climate change. It is constructed as a 0-100 indicator, where the country's commitment to environmental goals increases

with the score. The overall indicator is calculated from the weighted sum of four components: per capita GHG emissions (40% weighting), renewable energy (20% weighting), energy use (20% weighting), and climate policy (20% weighting), totaling 14 indicators. The rationale behind choosing these four components is that effective climate policy will influence energy use and renewable energy over a few years, ultimately reducing GHG emissions.

After data cleaning and matching, the final sample comprises deals originated between 2011 and 2018, resulting in 8,488 observations. These observations correspond to 1,951 unique deals granted by 185 distinct lenders to 556 unique borrowers headquartered in 33 countries. The borrowing firms operate in 56 two-digit SIC industries, corresponding to 11 industrial sectors, including the most carbon-intensive ones (oil, coal, gas, utilities, and materials).

2.3 Methodology and main variables

We run a fixed-effects panel regression analysis where the dependent variables are the cost, the amount, and the share of syndicated loans granted to polluting companies.

To account for the interlinkages of multiple risk drivers, we combine the measures of borrower pollution, the borrower's country's resilience to climate risk, and the binary variable "post-Paris Agreement", which constitutes the third prong of our CTR indicator. Our comprehensive measure of exposure to climate transition risk is, therefore, the following triple interaction:

$$\text{CO}_2 \text{ emissions}_{t,f,c} \times \text{CCPI}_{t,c} \times \text{Post}_t$$

where CO₂ emissions quantifies the total carbon emissions in thousands of tons for borrowing firm *f* in country *C* in year *t*, CCPI is the Germanwatch's Climate Change Performance Index of the borrower's home country in year *t*, and Post is a dummy variable taking the value of one after the signing of the Paris Agreement (years 2016 to 2018).

The intuition is that for each level of pollution, firms located in countries that are more environmentally conscious are more likely, since the Paris Agreement, to incur in sanctions and limitations designed to mitigate their carbon impact. This could affect firms financially and require expensive investments to adjust practices and business models. In turn, lenders should adjust their policy as an effect of higher transition risk, for example, by charging higher interest rates and/or allocating less credit to more exposed borrowers.

¹ Germanwatch provides measures for 57 countries and the E.U. (germanwatch.org)

We also investigate the non-linearity of banks' reactions to climate transition risk by looking at the cost and amount of credit to extremely vulnerable counterparties, namely highly polluting firms located in countries strongly committed to environmental issues. Our main explanatory variables become:

$$\text{Vulnerable}_{t,f} \times \text{High CCPI}_{t,c} \times \text{Post}_t$$

where vulnerable to transition risks are firms with CO₂ emissions above a given percentile in a specific year and High CCPI are countries with a climate index score above a given percentile in the index distribution in a given year. For both, the relevant thresholds are the 50th and the 75th percentiles of the distribution.

In investigating lending policies, we control for several time-varying and time-invariant factors at the loan, bank, firm, and country level that may influence bank lending policies. In particular, loan-level controls include the loan amount and maturity, the number of lead arrangers participating in the syndicate, as well as dummies for loan purpose and type, and the presence of covenants, performance pricing grid, and collateralization. Time-varying firm characteristics refer to borrowers' size, leverage, and profitability, all

lagged by one year. Bank-level variables control for size, capitalization, and profitability of individual banks (the lead arrangers). We also include bank fixed effects, so as to allow for time-invariant characteristics that may affect spreads and lending choices. To better control for peculiar characteristics on the demand side, we employ fixed effects for borrower industry as well as time-varying controls at the country level (namely, the GDP growth and the change in monetary policy rates). Moreover, we include year fixed effects to capture year-specific movements that may influence the corporate loan market and are common to all banks in the sample.

3. MAIN RESULTS

We obtain several findings.

First, we document a positive association between CO₂ emissions, loan prices, and loan supply over the entire time span considered. This suggests that banks were already mindful of their borrowers' environmental impact, as indicated by the higher interest rates applied to larger emitters, even before COP21. Simultaneously, credit to these borrowers has increased as CO₂ increased.



Second, the direction of the relationships between loan variables and CO₂ emissions reverse in the years following COP21, with both credit availability and loan prices decreasing as emissions increase. This indicates a shift in lending practices since the Paris Agreement, with banks granting less credit but at a lower price to larger emitters.

Furthermore, the relationship between loan variables and climate risk is non-linear and depends on both the climate vulnerability of the borrowers (proxied by high level of CO₂ emissions) and the climate resilience of the government in the borrowers' home country (proxied by high level of CCPI index). Specifically, we document a positive correlation between loan prices and borrowers' carbon emissions for highly vulnerable firms located in highly climate-resilient countries after COP21. These firms receive, on average, larger loan amount, but a lower share of loans after the Paris Agreement, suggesting a reallocation effect within the loan portfolio mix.

When we measure vulnerability not as firm-level CO₂ emissions, but by grouping borrowers based on the industry-level carbon intensity, we observe that the price effect of increased transition risk becomes stronger. Borrowers from more polluting industries headquartered in climate resilient countries are charged higher prices following the Paris Agreement. At the same time, banks have increased their exposure to these more polluting industries, not only in terms of the amount but also in the share of loans allocated to them, with no evidence of reallocation within the loan portfolio. These contrasting results underscore the importance of having detailed data that captures the climate sensitivity of bank exposures at different levels.

The baseline results concerning loan price and loan amount seem to be driven by European banks. Interestingly, we find no evidence that banks adhering to green standards are incorporating increasing climate transition risk in their lending practices differently from non-green banks.

4. CONCLUSION

We examine bank lending behavior in a context of increasing climate transition risks. By using a granular sample obtained by merging corporate, lender, and country information to syndicated loans data, we investigate two relevant dimensions for bank lending, namely loan pricing and supply. Our objective is to determine whether banks incorporate climate transition risks into loan pricing and whether they reduce credit (both in terms of loan amount and share of total loans) to borrowers that are more exposed to climate transition risk.

We provide a comprehensive measure of exposure to climate transition risk, considering three important risk drivers: the borrower's carbon emissions, a policy shock represented by the 2015 Paris Agreement, and climate resilience and policy stringency of the country in which borrowers are located.

After controlling for all these factors, we uncover that policy shocks, such as the Paris Agreement and government commitments to environmental issues, are important climate risk drivers that, when combined, amplify banks' perception of transition risk.

However, banks' responses to increased climate transition risk are neither uniform nor straightforward, and the relations among relevant variables are not linear. In terms of policy implications, our findings underscore the importance of comprehensively measuring firms' exposure to climate transition risk, considering both idiosyncratic and country-specific factors. Similarly, banks' exposure to climate-related risk needs to be assessed at both firm and industry levels, as evidence on banks' reactions to climate-related issues may vary depending on the proxy used.

Our findings do not support the hypothesis that banks labeled as "green" react to climate transition risk differently than non-green banks. This points to banks' greenwashing and suggests that not all initiatives promoted as environmentally friendly are equally effective.

More empirical evidence, supported by cleaner data on banks' and firms' exposure, would be helpful to clarify the role played by banks in the transition process, including whether any reallocation across firms and within industries has actually been taking place.

REFERENCES

BIS, 2021, "Climate-related risks drivers and their transmission channels," Basel Committee on Banking Supervision, <http://tinyurl.com/v23kxs55>

Bolton, P., and M. Kacperczyk, 2021, "Do investors care about carbon risk?" *Journal of Financial Economics*, 142:2, 517-549

Bruno, B., and S. Lombini, 2023, "Climate transition risk and bank lending," *Journal of Financial Research* 46:1, 59-106

Delis, M., K. de Greiff, M. Iosifidi, and S. Ongena, 2023, "Being stranded with fossil fuel reserves? Climate policy risk and the pricing of bank loans," *Financial Markets, Institutions and Instruments*, 1-27, <http://tinyurl.com/5freez6d>

Donefer, C., 2023, "State ESG laws in 2023: the landscape fractures," Thomson Reuters, May 31, <http://tinyurl.com/yj6tx8py>

HOW FINANCIAL SECTOR LEADERSHIP SHAPES SUSTAINABLE FINANCE AS A TRANSFORMATIVE OPPORTUNITY: THE CASE OF THE SWISS STEWARDSHIP CODE

AURÉLIA FÄH | Senior Sustainability Expert, Asset Management Association Switzerland (AMAS)

ABSTRACT

This article explores the pivotal role that financial services play in advancing sustainable finance, with a focus on the Swiss Stewardship Code published in October 2023 as a case study. It highlights the financial services sector's inherent bias toward recognizing and capitalizing on the transformative opportunities presented by sustainable finance, emphasizing long-term value creation, risk management, and innovation. It contrasts market-based and regulatory approaches to sustainability, showing Switzerland's preference for market- and principle-based approaches. The Swiss Stewardship Code, developed by the Asset Management Association of Switzerland and Swiss Sustainable Finance, is presented as a model for effective stewardship in sustainable investing. The article argues that this approach, emphasizing collaboration, innovation, and a proactive stance towards sustainability, not only combats greenwashing but also aligns financial flows with sustainability goals, underscoring the financial services sector's critical role in driving sustainable economic, social, and environmental outcomes.

1. INTRODUCTION

In recent years, sustainable finance has evolved from a niche field to a critical component of a broader strategy to transition towards a more sustainable economy and to align global financial flows with sustainability goals. Such ambitious objectives require the collaboration of all key stakeholders, ranging from corporates, financial players, consumers, and policymakers. Market-based approaches have historically been critical in shaping sustainable finance practices over the past decades. More recently, regulatory initiatives have flourished around the world to create the framework and the conditions for the integration of environmental, social, and governance (ESG) factors into financial services. Depending on the jurisdictions, they mostly limit themselves to the prevention

of the risks associated with the negative aspects of sustainable finance, including greenwashing. On the other hand, market-based initiatives often focus on the transformative potential and opportunity associated with sustainability.

Switzerland follows a market- and principle-based approach. The recent publication of the Swiss Stewardship Code in October 2023 by industry associations demonstrates the ambition of the financial services sector to keep leading the way and creates the necessary standards placing sustainability as an opportunity for the Swiss financial industry.

This article seeks to explore how financial sector leadership advances the transformative opportunity of sustainable finance by diving into the recent introduction of the Swiss Stewardship Code.

2. SEIZING THE TRANSFORMATIVE OPPORTUNITY OF SUSTAINABLE FINANCE

When it comes to seizing the transformative opportunity of sustainable finance, the financial services sector proves to be better equipped than other relevant stakeholders.

2.1. Financial services sector's bias toward positive and long-term value creation

The financial services sector possesses an inherent bias towards recognizing and capitalizing on the transformative opportunities presented by sustainable finance. This bias stems from several factors:

- **Risk management perspective:** financial institutions recognize the materiality of ESG factors in assessing risk. As sustainability issues such as climate change, resource scarcity, and social inequality become more prominent, financial institutions understand that integrating ESG considerations into their decision-making processes is essential for long-term risk management and value preservation.
- **Long-term value creation:** sustainable finance offers opportunities for long-term value creation and resilience. Investments in sustainable projects and businesses not only generate financial returns but also contribute to environmental protection, social development, and economic growth. Financial institutions that prioritize sustainability are well-positioned to create lasting value for their stakeholders and society as a whole.
- **Client demand and investor preferences:** there is a growing demand from clients and investors for sustainable finance products and services. As awareness of sustainability issues increases, individuals and institutions are seeking investment opportunities that align with their values and contribute to positive environmental and social outcomes. Financial institutions are responding to this demand by offering a wide range of sustainable investment options, thereby capitalizing on the transformative opportunity presented by sustainable finance.
- **Market opportunities:** the transition to a sustainable economy presents significant market opportunities for financial institutions. Investments in renewable energy, clean technology, sustainable infrastructure, and other environmentally and socially responsible sectors offer the potential for attractive returns while also addressing

pressing sustainability challenges. Recognizing these opportunities, financial institutions are increasingly allocating capital towards sustainable finance initiatives to capture market share and drive innovation.

- **Financial innovation:** sustainable finance drives financial innovation by creating new investment opportunities, products, and services that integrate ESG considerations. Innovations, such as green bonds, impact investing, and sustainability-linked loans, mobilize capital towards sustainable projects and businesses, unlocking new sources of financing and stimulating economic growth.
- **Reputational and brand considerations:** financial institutions recognize the importance of sustainability in building and maintaining their reputation and brand value. Embracing sustainable finance practices enhances their credibility, attracts clients and investors, and strengthens relationships with stakeholders. By demonstrating a commitment to sustainability, financial institutions can differentiate themselves in the market and gain a competitive advantage.

2.2. Policymakers' inherent focus on risk mitigation and investor protection

Policymakers are primarily focused on addressing the negative consequences of sustainable finance, such as greenwashing, for several reasons:

- **Risk mitigation:** regulators have a responsibility to protect investors and consumers from misleading or deceptive practices, including greenwashing. By focusing on the negative aspects of sustainable finance, regulators aim to mitigate the risks associated with false or exaggerated claims of environmental or social responsibility.
- **Market integrity:** ensuring market integrity is essential for maintaining trust and confidence in the financial system. Regulators seek to prevent greenwashing to safeguard the integrity of sustainable finance markets and prevent market manipulation or fraud that could undermine investor confidence and market stability.
- **Investor protection:** regulators prioritize investor protection by requiring transparency and disclosure of material information related to ESG factors. By addressing greenwashing and ensuring accurate and reliable information, regulators aim to empower investors to make informed decisions and protect them from potential harm or financial losses.

- **Regulatory compliance:** regulators enforce laws and regulations related to sustainable finance to ensure compliance with legal standards and prevent violations of consumer protection and securities laws. Focusing on the negative aspects, such as greenwashing, helps regulators identify and address instances of non-compliance and hold financial institutions accountable for their actions.
- **Market stability:** greenwashing and other misleading practices in sustainable finance can create market distortions and undermine the efficient allocation of capital. Regulators aim to maintain market stability by addressing greenwashing and promoting transparency, integrity, and accountability in sustainable finance markets.
- **Public trust and confidence:** governments and regulators recognize the importance of public trust and confidence in the financial system. Addressing greenwashing and promoting integrity in sustainable finance markets are essential for maintaining public trust and confidence in the credibility and effectiveness of sustainability initiatives.

Overall, regulators and governments tend to address the mitigation of the negative aspects of sustainable finance, such as greenwashing, to protect investors, maintain market integrity, ensure regulatory compliance, promote market stability, uphold public trust, and advance long-term sustainability goals. By addressing greenwashing and other misleading practices, regulators aim to foster a more transparent, responsible, and effective sustainable financial industry that delivers positive environmental, social, and economic impact.

The financial services sector is, on the other hand, strongly equipped to seize the transformative opportunity of sustainable finance because of its agility, innovation capacity, and direct influence on investment flows. Financial institutions can quickly adapt to market trends, integrate ESG criteria into their investment decisions, and develop new financial products that support sustainable development goals. This agility allows the financial services sector to respond promptly to investor demands for sustainable options, driving change efficiently and effectively across economies.

The prominence of the financial services sector in embracing sustainable finance as an opportunity for the financial industry proved particularly true in the Swiss context.

3. THE PIVOTAL ROLE OF THE FINANCIAL SERVICES SECTOR IN ADVANCING SUSTAINABLE FINANCE PRACTICES IN SWITZERLAND

In Switzerland, the financial services sector has been playing a critical role in shaping and advancing sustainable finance practices. Through a combination of self-regulatory initiatives and private-led best practices, such market-driven approaches offer an effective and ambitious alternative to fully-fledged regulatory approaches undertaken by similar jurisdictions such as the E.U.

3.1. Rationale for a private sector-led approach in Switzerland

By way of background, Switzerland is particularly favorable to a **market-based approach**. Such an approach is rooted in the country's political and economic history, as well as its commitment to principles such as liberalism, free enterprise, economic freedom, and individual responsibility. By fostering a dynamic and competitive market environment, Switzerland aims to promote innovation, growth, and prosperity while maintaining social cohesion and environmental sustainability.

Additionally, the **subsidiarity principle** is a guiding concept in Swiss governance, emphasizing that decisions should be made at the most immediate or local level, only involving higher levels of government if necessary. This principle supports a market-based approach to the economy, where the market and private entities play a significant role in economic activities, and government intervention is minimized.

3.2. The role of self-regulations

In Switzerland, self-regulation in finance is a significant component of the regulatory framework, complementing formal legislation and oversight by regulatory authorities. Financial institutions and industry associations are deemed most appropriate to develop and enforce their own sets of rules and standards to promote ethical behavior, transparency, and efficiency within the market. These self-regulatory organizations cover all financial services sectors, including banking, asset management, and insurance, aiming to uphold the integrity and stability of Switzerland's financial system while fostering innovation and competitiveness. The Swiss Financial Market Supervisory Authority (FINMA) supports this model and recognizes three types of self-regulation: voluntary self-regulation, self-regulation recognized as a minimum

standard, and compulsory self-regulation. This framework enables the financial services sector to develop standards in close collaboration with experts, ensuring market relevance and broad acceptance. Self-regulation is instrumental in complementing and detailing key areas of state regulation, with FINMA having the authority to recognize and enforce self-regulatory guidelines as minimum standards. This ensures that not only members of self-regulatory organizations but also other sector participants adhere to these guidelines.

When it comes to sustainability, financial industry associations have developed their self-regulations over the past two to three years:

- The **Asset Management Association of Switzerland (AMAS)** has developed a principle-based self-regulation for sustainable asset management released in September 2022 and effective since September 2023. Its framework for sustainable asset management lays down the organizational requirements for financial institutions, as well as for product design and disclosures to investors, to prevent and combat greenwashing by enhancing the quality of collectively managed sustainable assets through binding standards, while improving transparency through comprehensive documentation and reporting obligations. With its explicit references to both institutional and product levels, the AMAS self-regulation dovetails with the self-regulation process of client advisory that the Swiss Bankers Association has introduced.
- The **Swiss Bankers Association (SBA)** elaborated a principle-based self-regulation for the providers of financial services on the integration of ESG preferences and ESG risks into investment advice, portfolio management, and mortgage advice, which was first published in June 2022 and is effective since January 2023.
- The **Swiss Association of Pension Funds (ASIP)** published in December 2022 a standard for ESG reporting for Swiss Pension funds that came into force in the financial year 2023.
- As we write, the **Swiss Insurance Association (SIA)** is working with its members to elaborate a self-regulation to be published in the coming months.

The elaboration of self-regulation is conducted through the effective collaboration of financial stakeholders and led by their respective industry associations. Sustainable finance-

related self-regulations in Switzerland support the objectives of the Swiss authorities and their sustainable finance strategy. In particular, the Federal Council published in December 2022 a position focusing on the prevention of greenwashing, which aligns with the objectives advanced by the industry self-regulations. With self-regulations already published, the financial industry proactively took the necessary steps on its own to prevent greenwashing, foster transparency, and safeguard the credibility of the Swiss financial center.

3.3. The importance of other private-led initiatives

Beyond self-regulatory mechanisms, numerous private-led initiatives stand at the forefront of advancing sustainable finance in Switzerland. By their nature, those initiatives usually go beyond the mitigation of the negative aspects of sustainable finance and focus instead on the opportunity inherently associated to sustainability.

- **International best practices:** historically, the private sector stood as the historical lever to advance sustainability best-practices globally. The concept of integrating sustainability characteristics into finance was first advanced in 1992 during the Earth Summit in Rio de Janeiro. The transformation of private finance was recognized as essential for achieving sustainable development and led to the creation of the U.N. Environment Program Finance Initiative (UNEP FI), a partnership between UNEP and the global financial services sector. Further standards and metrics, such as the Global Reporting Initiative (GRI) in 1997 or the Principle for Responsible Investment (PRI) in 2006, were subsequently developed by, or in close collaboration with, the financial services sector.
- **Net-zero initiatives:** more recently, the Swiss financial services sector actively joined international net-zero alliances to combat climate change and align with the goals of the Paris Agreement. This commitment is evident across banking, insurance, and asset management sectors, with significant participation in Global Financial Alliances Net Zero (GFANZ) related initiatives, including the Net Zero Asset Managers initiative, Net-Zero Banking Alliance, and Net-Zero Insurance Alliance.¹ Those initiatives proved to be particularly effective. In the case of the Net Zero Asset Management (NZAM) initiative, AMAS reports that as of September 2023, a total amount of CHF 628 billion (approximately U.S.\$713 billion) of

¹ PwC, 2022, "Setting sail for a carbon-neutral future: Net Zero Insights 2022," <https://tinyurl.com/yecjf75e>

AMAS' members' AuM are currently managed in line with net-zero, which represents an increase of 18% compared to December 2022 levels. These private-led efforts are supported by the Swiss government and aim to standardize credible climate targets and increase sustainable finance's role in achieving net-zero emissions by 2050.

- **Swiss Stewardship Code:** another concrete example of a recent and private-led initiative building on the above, and paving the way to advance sustainable finance as a transformative opportunity, is the Swiss Stewardship Code. The Code elaborated by industry associations will be the subject of a particular case study in the final section.

3.4. Pros and cons of the Swiss private sector-led approach to sustainable finance

The key advantages of a market-led approach in a Swiss context mirror elements highlighted in Section 2, above. Industry-led self-regulations, as well as private-led initiatives, have the merit of having been developed by the industry for the industry, which make them particularly effective and fit for purpose. In a fast-moving field, such as sustainable finance, they additionally present the fantastic advantage of being agile and flexible. Those approaches were indeed elaborated in a six to nine months' timeframe and can easily and regularly be amended to reflect the latest international best practices when the initiators deem suitable and appropriate.

On the other hand, commonly referred to drawbacks of such an approach often include the lack of enforcement of self-regulations and private-led initiatives, even though self-regulations are actually binding on the members of the industry associations represented. By and large, industry associations include more than two-thirds of the market represented (in terms of assets under management, for example, for AMAS). When it comes to other initiatives, such as net-zero alliances, market competitiveness ultimately encourages market players to apply ambitious standards as highlighted in Section 2.

4. THE CASE OF STEWARDSHIP – THE SWISS STEWARDSHIP CODE

In October 2023, the Asset Management Association of Switzerland (AMAS) together with Swiss Sustainable Finance (SSF) published the Swiss Stewardship Code.² The Code exemplifies point to point how a market-led approach contributes to tackling the most transformative aspects of

sustainable finance. The Code sets forth principles for effective stewardship applicable across the industry, encompassing both asset managers and owners. It was developed through a collaborative effort involving a broad spectrum of investors, including both asset owners and managers, in addition to service providers.

4.1. Stewardship as one of the most critical approaches to achieving positive change

Investment stewardship is a responsible investment approach by which investors collaborate and interact with investee entities with the aim of generating long-term financial, environmental, and societal value. This investment approach has always been used in the financial services sector and the real economy. In recent years, however, stewardship, and more specifically voting and engagement, has become increasingly important as investors have started to expand their goals to encompass the contribution to positive change in the economy, in society, and for the environment. Amongst the different sustainable investment approaches, stewardship proved particularly effective in achieving positive impact, in tackling sustainability-related challenges, and in addressing sustainability-related risks. As opposed to other sustainable investment approaches, such as exclusion, for example, stewardship aims at collaborating with investee companies to lead them through the necessary transformations required to reach positive and long-term sustainable outcomes. In the case of a climate goal, for example, the investment approach directly aims at decarbonizing the real economy through active dialogue with a company. On the other hand, an investment approach based on exclusion would artificially decarbonize an investment portfolio without leading to any change in the real economy, where the excluded company may access the financing from a less stringent type of investor.

4.2. Effective stewardship through the application of nine ambitious principles

The Code has several objectives. First, it aims at elaborating standards defining stewardship as a sustainable investment approach that proves effective in achieving positive impact on key sustainability-related challenges. Integrating stewardship into the investment processes of the Swiss investment industry promotes a more sustainable and value-adding economy and helps to increase long-term returns for investors, adjusted for sustainability risks. Second, it aims to provide a level-playing field for Swiss stakeholders involved in stewardship

² AMAS and SSF, 2023, "Swiss Stewardship Code," <https://tinyurl.com/mr3xbwcz>

activities. This level-playing field lays the foundation for higher transparency and better comparability of stewardship practices. The need for transparency in stewardship activities was also highlighted as a key area of action by the Federal Council in its 2022-2025 strategy on sustainable finance.³

While being focused on the Swiss investors' practices, the Code builds on the extensive local and international experience of AMAS and SSF members. Global standards and international best practices, such as the Global Stewardship Principles of the International Corporate Governance Network (ICGN), the U.N. Principles for Responsible Investing (PRI), and the U.K. Stewardship Code, represent international benchmarks for stewardship activities by investors.

The Swiss Stewardship Code comprises nine stewardship principles and describes the key elements for effective and successful implementation. The key principles of the Code focus on recommendations related to those two critical means by which stewardship is commonly achieved: voting and engagement in an active dialogue with the companies. When it comes to engagement, for example, the Code emphasizes the importance of engaging in an active dialogue at different levels. An active dialogue can indeed be conducted by an investor on an individual basis or can also be conducted collaboratively with other investors or service providers to heighten engagement outcomes. Beyond their engagement with investee companies, investors may also decide to engage in an active dialogue with relevant public stakeholders and policymakers. The latter aspect is a unique aspect that the U.K. Stewardship Code, for example, does not provide for. A critical principle of the Code addresses the importance of defining the conditions under which an engagement is considered to be failing, as well as the conditions under which an escalation may be triggered. In the latter case, relevant escalation steps may go as far as divestment. The code also tackles key elements related to monitoring and reporting, as well as the management of conflict of interest and the delegation of stewardship activities to a service provider.

4.3. An agile and market-driven approach

Elaborated for practitioners by practitioners, the Swiss Stewardship Code was developed by AMAS and SSF, with the expertise of their members' specialists ranging from

asset managers, asset owners, and service providers. The Code acknowledges the diverse nature of asset owners, asset managers, and service providers, each varying in size, business model, and investment approach, leading them to exercise stewardship in different ways. This code represents a groundbreaking step towards unifying and enhancing stewardship activities within the country. Collectively, asset owners, asset managers, and service providers form an intricate web of responsible investing, each with a unique role and responsibility. As the Swiss Stewardship Code prepares to take center stage, it is this collaborative ecosystem of investors that holds the power to drive positive change in Switzerland's financial industry. The commitment of the industry is, therefore, expected to be stronger than a regulatory-led initiative. Additionally, and content-wise, it is also expected to be more ambitious. One of the key differences between the U.K. Stewardship Code and the Swiss Stewardship Code from a content perspective is that the Swiss Stewardship Code's principles include public policy engagement. The latter is not to be found in the U.K. Code since it was developed by the regulator, the Financial Conduct Authority (FCA).

5. CONCLUSION

The financial services sector plays a pivotal role in shaping and advancing sustainable finance practices. Through the market-led and principle-based approach exemplified by the Swiss Stewardship Code, Switzerland's financial services sector demonstrates an effective alternative to fully regulatory models, leveraging self-regulations, innovation, and international collaboration to address sustainability challenges. The Swiss case underscores the potential of financial services sector leadership to drive transformative opportunities in sustainable finance, highlighting the importance of collaboration, innovation, and a proactive stance towards sustainability. This approach not only fosters transparency and combats greenwashing but also aligns financial flows with long-term sustainability goals, creating value for the economy, society, and the environment. The Swiss Stewardship Code, with its focus on responsible investment and stewardship, serves as a blueprint for engaging the financial services sector, emphasizing the sector's pivotal role in achieving a sustainable future.

³ Federal Council report, 2022, "Sustainable finance in Switzerland: areas for action for a leading sustainable financial center, 2022-2025," <https://tinyurl.com/28sv9dj5>

GOVERNANCE OF CORPORATES



126 Cycles in private equity markets

Michel Degosciu, CEO, LPX AG

Karl Schmedders, Professor of Finance, IMD

Maximilian Werner, Associate Director and Research Fellow, IMD

134 Higher capital requirements on banks: Are they worth it?

Josef Schroth, Research Advisor, Financial Stability Department, Bank of Canada

140 From pattern recognition to decision-making frameworks: Mental models as a game-changer for preventing fraud

Lamia Irfan, Applied Research Lead, Innovation Design Labs, Capco

148 Global financial order at a crossroads: Do CBDCs lead to Balkanization or harmonization?

Cheng-Yun (CY) Tsang, Associate Professor and Executive Group Member (Industry Partnership), Centre for Commercial Law and Regulatory Studies (CLARS), Monash University Faculty of Law (Monash Law)

Ping-Kuei Chen, Associate Professor, Department of Diplomacy, National Chengchi University

158 Artificial intelligence in financial services

Charles Kerrigan, Partner, CMS

Antonia Bain, Lawyer, CMS

CYCLES IN PRIVATE EQUITY MARKETS¹

MICHEL DEGOSCIU | CEO, LPX AG

KARL SCHMEDDERS | Professor of Finance, IMD

MAXIMILIAN WERNER | Associate Director and Research Fellow, IMD

ABSTRACT

In this study, we analyze three decades of return data from listed private equity (LPE) companies, focusing on the return averages and volatilities of two notable market indices and comparing them to a global equity index. Our findings indicate that LPE has generated higher average returns, commensurate with its higher volatility, in comparison to the global index. Additionally, we observe that, on average, LPE companies have traded at a discount to their book values since the Great Financial Crisis. Importantly, this discount exhibits a strong negative correlation with an indicator of macro-financial stress, which emerges as a predictive factor for LPE market performance.

1. INTRODUCTION

In recent years, the aftermath of the COVID-19 pandemic and the ongoing conflict in Ukraine have profoundly reshaped the global economic landscape. Nations worldwide are grappling with a resurgence of inflation, a challenge that had remained largely dormant for decades. The U.S., the E.U., Canada, Australia, and Japan, among other countries, have all experienced consumer price index increases² not seen in over thirty years. This significant surge in inflation across these major economies has highlighted substantial economic shifts, manifesting in a widespread and impactful rise in the cost of goods and services. This inflationary wave, fueled by external shocks and the strategic responses of governments and central banks, prompted a notable increase in interest rates throughout 2022 and 2023. Figure 1 presents the respective time series for monthly inflation within the eurozone and its monthly risk-free rate (derived from German treasury bills).³

Unsurprisingly, this economic environment has posed significant challenges for investors, who have been navigating the repercussions of these inflationary pressures for asset

values, interest rates, and investment strategies. This has marked a period of recalibration and of heightened uncertainty in global financial markets. A 2023 survey of global institutional investors revealed that this macro-financial regime shift has been a top priority of investors.⁴ The survey reports that 80% of participating investors agreed “that the world is dramatically changing and that portfolios must evolve to keep pace,” 56% recognized “that the current environment is unlike any they’ve seen in their careers,” and 64% expected their “inflation mitigation strategies” to have a duration of two years or more. The survey further documents that, as investors have had to navigate the complexities of a different economic climate, a growing inclination toward diversifying portfolios with private assets has emerged. A striking 72% of survey participants planned “to increase their private markets allocation over the next five years.” This striking proportion naturally leads us to ask why so many institutional investors want to increase their exposure to private markets during a time of heightened economic uncertainty.

Among private market investments, “listed private equity”

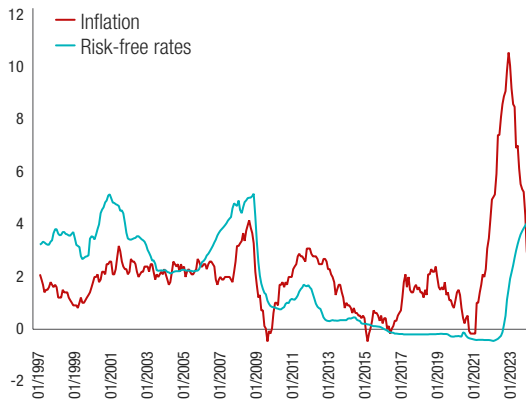
¹ We are very grateful to Jonas Vogt for helpful discussions and to Dave Brooks for outstanding editorial support on previous versions of this manuscript.

² <http://tinyurl.com/ycy5vk7u>

³ Throughout the present paper, the data for the monthly risk-free rate is computed from OECD data on German short-term interest rates. The combined and transformed data covers the period 12/31/1993 to 12/29/2023. The data was taken from the OECD’s data portal. See Section 2.2 for further details.

⁴ <http://tinyurl.com/35ysk6yb>

Figure 1: Monthly inflation and the risk-free rate in the eurozone



Monthly inflation and the risk-free rate in the eurozone for the period 01/01/1997 to 12/29/2023. Inflation refers to the HICP – Overall index (ICP.M.U2.N.000000.4.ANR), published on the European Central Bank (ECB) data portal; last accessed January 5, 2024. Both rates are given as percentages.

(LPE) stands out as a particularly intriguing option due to its unique blend of private equity's potential returns and the liquidity of public markets. In this article, we delve into a targeted examination of LPE investments. Specifically, we analyze the interplay between future returns, price-to-book ratios, and the landscape of macro-financial uncertainty. Our investigation posits that periods of macro-financial distress can often lead to a structural underestimation of (listed) private equity's value, presenting savvy investors with opportunities for substantial gains.

2. LISTED PRIVATE EQUITY (LPE)

Private equity (PE) refers to investment funds that directly invest in private companies or engage in buyouts of public companies, resulting in these companies delisting from public stock exchanges. These funds are managed by professional investment firms and aim to create value through strategic improvements, operational efficiencies, and leveraging industry expertise. PE investments are typically characterized by long investment horizons and active management, with the goal of exiting these investments through sales or public offerings at a significant profit.

Under the broader umbrella of PE, a specific subgroup known as LPE exists. LPE firms are those PE entities that are themselves publicly traded on a stock exchange, offering investors the unique opportunity to engage with PE investments through

publicly accessible shares. This arrangement combines the investment strategies of PE – such as direct investments in private companies, leveraged buyouts, and venture capital – with the liquidity and accessibility of public markets. LPE allows individual and institutional investors to partake of the potential returns of PE investments without the typical barriers to entry, such as high minimum investment thresholds and long-term capital commitments.

The common challenge within the realm of PE is the notable scarcity of accessible, reliable data. Transactions in PE typically involve unlisted companies, rendering the details of these deals largely opaque and seldom observable through hard, quantitative data. This lack of transparency can significantly hinder the ability of investors to perform thorough due diligence, accurately assess the value and potential of investments, and benchmark performance against relevant indices or competitors.

In contrast, LPE offers a compelling advantage in this regard. Being publicly traded entities, LPE firms are obligated to adhere to the regulatory requirements of stock exchanges, which mandate regular financial reporting and disclosure of material information. This transparency ensures that a wealth of data is available to investors, encompassing financial performance, investment strategies, and market positioning. Such information not only facilitates a more informed investment decision-making process, it also enables ongoing monitoring and analysis of the investment's performance. Consequently, LPE can serve as a bridge for investors seeking exposure to the PE sector's potential rewards, coupled with the transparency, liquidity, and data availability characteristic of public markets. This duality underscores the unique value proposition of LPE, marrying the growth and return potential of PE investments with the advantages of public market accessibility.

2.1 Data on LPE

For our analysis of LPE returns, we use two LPE indices provided by LPX AG, a Zurich-based provider of financial market data. The first index, the “LPX50 Listed Private Equity Index TR” (LPX50), offers a diversified representation across various dimensions, including geographic regions, PE investment styles, financing methods, and vintage years, thereby ensuring a comprehensive reflection of the LPE market. For our return analysis in this article, we use (EUR-based) month-end index values of LPX50 from December

1993 until December 2023. The second index, the “LPX Buyout Listed Private Equity Index TR” (LPXBO), is designed to represent the global performance of those LPE companies that pursue a buyout PE investment strategy. Buyout funds specialize in acquiring controlling interests in companies with the aim of increasing their value over time before eventually selling those companies for a profit. The LPXBO comprises the 30 most highly capitalized and liquid LPE companies, again diversified across regions, financing styles, and vintages. For the LPXBO we also use (EUR-based) closing monthly index values from December 1993 until December 2023.

The calculation of LPX index levels requires only two simple components: the share prices of the LPE firms included in the index and their relative index weights. However, understanding the fundamental value of an LPE firm requires navigating a more complex aspect. The share price of an LPE firm might not accurately reflect the total value of its investments in private companies, primarily because these investments lack publicly observable prices. Instead, the valuation of these private investments often relies on their book values, which are estimated figures that attempt to quantify the worth of the LPE firm’s investments. And the sum of these book values provides an estimate of the LPE firm’s book value.

Benjamin Graham’s insightful observation to Warren Buffet,⁵ “Price is what you pay; value is what you get,” eloquently highlights the difference between the market price and the intrinsic value (book value) within the context of LPE firms. It is important to note that there is typically a discrepancy between the sum of an LPE firm’s investment book values and its market capitalization. This difference underscores the challenge investors face in assessing the true value of LPE firms, as it requires looking beyond share prices to understand the underlying stocks’ estimated worth.

Building on the distinction between the market price and the intrinsic value of LPE firms, it is pivotal for investors to explore the concept of premia and discounts in their market valuation. A market price trading at a premium indicates that the market value of an LPE firm exceeds the aggregate book value of its investments, suggesting that investors are willing to pay more than the estimated value of the underlying assets. This premium could be attributed to factors such as the management team’s track record, anticipated growth of the portfolio companies, or the firm’s strategic positioning within a high-growth sector.

Conversely, a market price trading at a discount to the aggregate book value of its investments implies that the market values the LPE firm less than it does the sum of its parts. This discount could arise from various concerns, including market skepticism about the valuation of the underlying investments, potential liquidity issues, or broader economic uncertainties impacting investor sentiment. Discounts offer an intriguing opportunity for investors who believe that the market has undervalued the LPE firm’s portfolio, presenting a chance to invest in the firm’s assets at a price lower than their perceived intrinsic value.

In our data analysis, we enhance the evaluation of the two LPE indices by incorporating their respective price-to-book ratios.⁶ To specifically gauge the premium or discount at which each index is trading, we employ the following premium/discount (PD) indicator:

$$\frac{(\text{market price} - \text{book value})}{\text{book value}} = \frac{\text{market price}}{\text{book value}} - 1$$

This calculation clearly delineates the relationship between market capitalization and book value, providing a quantifiable measure of valuation sentiment. We have access to monthly data on the respective indicator for LPX50 and LPXBO from December 2002 until December 2023.

When the PD indicator yields a positive value, it signifies that the market capitalization of the index surpasses its book value, indicating that, on aggregate, the stocks within the index are trading at a premium. Conversely, a negative indicator value suggests that the market capitalization is less than the book value, denoting that, collectively, the stocks are trading at a discount. This methodology provides insights into current market perception, revealing whether investors are valuing the index components as worth more or less than their estimated net assets.

2.2 Additional data

To gauge the returns of the global stock market, we use the MSCI World Net TR Index (MSCI in the remainder of the article) on its EUR basis. This comprehensive index represents the performance of publicly listed large- and mid-cap companies across 23 developed market economies. The index captures about 85% of the free-float adjusted market

⁵ <http://tinyurl.com/yckbbdp9>

⁶ The data on the indexed book values for LPX50 and LPXBO is from LPX AG’s database.

capitalization in each participating country. We transform OECD data⁷ on German treasury bill rates to obtain a measure for the monthly risk-free rate in Europe. Our data on the market index and the risk-free rate covers the 360 months from January 1994 until December 2023.

In our analysis, we also employ an indicator of contemporaneous stress in the financial system. The Composite Indicator of Systemic Stress (CISS) is a financial stability indicator developed by the ECB to measure the systemic stress levels within the financial system of the eurozone.⁸ The CISS combines information from various financial markets – including equity markets, bond markets, foreign exchange markets, money markets, and financial intermediaries – to provide a comprehensive view of systemic stress. It is designed to capture the interconnectedness of industries and markets and the potential for stress in one market or sector to spill over into others, thereby affecting the financial system's stability. By aggregating these various indicators, the CISS offers a single, continuous measure of systemic stress in real time. We make use of CISS data for the 252 months from January 2003 until December 2023.

3. ANALYSIS OF LPE RETURNS

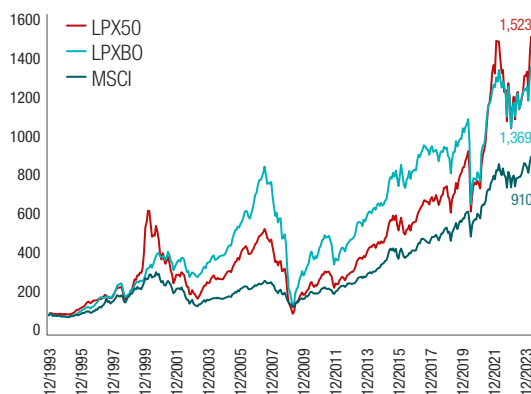
Figure 2 illustrates the cumulative monthly returns of LPX50 and LPXBO from December 31, 1993 to December 29, 2023, with the MSCI serving as a comparative benchmark. We mention several initial observations based on simple visual inspection. During the 30-year period, LPX50 and LPXBO significantly outperformed the MSCI benchmark, with absolute returns exceeding the benchmark by 67.4% and 50.4% respectively. Furthermore, both indices exhibit higher volatility compared to MSCI. This increased volatility is reflected in periods of significant outperformance followed by pronounced market corrections during times of economic downturn. Notably, major events, such as the bursting of the internet bubble in the early 2000s, the Great Financial Crisis (GFC) of 2007–2009, and the onset of the COVID-19 pandemic in early 2020, are distinctly visible in the trend lines. These events underscore the LPE indices' sensitivity to market dynamics, illustrating their potential for both higher rewards and higher risks.

3.1 Return statistics

We present some key summary statistics underlying our visual observations. Table 1 offers a closer look at the returns of LPX50, LPXBO, and MSCI indices, along with the risk-free rate. We observe that the average annualized (geometric) returns for LPX50 and LPXBO stand at 9.50% and 9.11% respectively, thereby notably outperforming MSCI's average annualized return of 7.64% during the past 30 years. This superior return performance of the LPE indices compared to MSCI underscores a possible reason for the attractiveness of this asset class among some groups of investors. Our second observation, the notably higher volatility of the LPE indices, is substantiated by their standard deviations (STD): 22.82% for LPX50 and 20.70 percent for LPXBO compared to 14.76% for MSCI. These quantitative results confirm the visual assessment of larger volatility in LPE markets.

A capital asset pricing model (CAPM) regression (based on data with 360 monthly excess returns) provides beta values of 1.26 for LPX50 and 1.07 for LPXBO. While both LPE indices exhibit a positive alpha, these are not statistically significant. The regressions yield R-squared values of 67% for LPX50 and 59% for LPXBO. For the LPE indices, which might be expected to have a higher component of specific (unsystematic) risk

Figure 2: Cumulative returns of LPX50, LPXBO, and MSCI



Cumulative monthly returns for LPX50, LPXBO, and MSCI for the period 12/31/1993 to 12/29/2023. The time series are normalized to the value of 100 on their starting date.

⁷ We take German treasury bill data from the OECD data portal. More precisely, we take the values for Germany of the OECD (2024) "short-term interest rates" (indicator) for the period 31/12/1993 to 30/11/2023. The missing data point for December 2023 is taken from the OECD (2024) "short-term interest rates forecast" (indicator) as the Q4 2023 forecast to complete the period December 1993 to December 2023. All data is transformed into a monthly time series.

⁸ The CISS (CISS.D.U2.Z0Z.4FEC.SS_CIN.IDX) data is from the ECB data portal. We took the NEW CISS series version instead of the original CISS and use the term CISS for simplicity. See Holló, D., M. Kremer, and M. Do Luca, 2012, "CISS – a composite indicator of systemic stress in the financial system," ECB working paper no. 1426, <http://tinyurl.com/2uzrcbc9>

Table 1: Return statistics

RETURN STATISTICS OVER 30 YEARS						
	RETURN	STD	BETA	SHARPE	SORTINO	TREYNOR
Private equity						
LPX50	9.50	22.82	1.26	0.43	0.64	0.077
LPXBO	9.11	20.70	1.07	0.44	0.61	0.084
Benchmarks						
MSCI	7.64	14.76		0.44	0.66	
Risk-free rate	1.97	0.54				
RETURN STATISTICS OVER 10 YEARS						
Private equity						
LPX50	12.98	20.85	1.35	0.69	1.03	0.106
LPXBO	8.09	19.39	1.23	0.50	0.70	0.078
Benchmarks						
MSCI	11.01	13.94		0.81	1.29	
Risk-free rate	0.16	0.33				

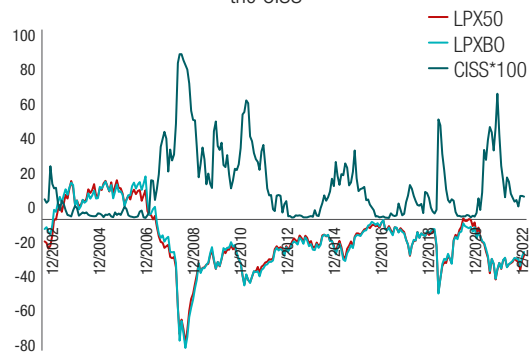
The reported figures are calculated from the 360/120 observations of the monthly returns for LPX50, LPXBO, MSCI, and the risk-free rate. All numbers, except for the beta values, are calculated with monthly data and then annualized using the standard annualization formulas and scaling factors. Averages and standard deviations are given as percentages. The ratios for the risk–return trade-offs are reported as decimals.

due to the nature of PE investments, these R-squared values suggest a stronger than expected correlation with the broader market. This result suggests that despite the PE nature of the LPE indices, the listed entities' returns are still significantly driven by market factors.

To further explore the risk–return trade-off, we report the Sharpe, Sortino, and Treynor ratios in Table 1. Notably, the (annualized) Sharpe ratios for all three indices are remarkably similar, suggesting that the higher returns associated with the LPE indices are proportionate to their increased volatilities. Similarly, the (annualized) Sortino ratios (with the reference point 0) are also close. In other words, the risk–return trade-off for the LPE indices aligns closely with that of MSCI. For completeness, we also report summary statistics for the most recent decade, from December 31, 2013 to December 29, 2023, in the bottom half of Table 1. This period was notably marked by the COVID-19 pandemic beginning in 2020 and the onset of the Russia–Ukraine war in 2022. These events significantly impacted financial markets, leading to observable changes in market volatility and trading volumes, as depicted in Figure 1.

We note that the average annualized (geometric) returns for LPX50 and LPXBO are 12.98% and 8.09%, respectively. This reveals that LPX50's return surpassed MSCI's average of 11.01%, whereas LPXBO's return did not. However, when

considering risk-adjusted performance, both LPE indices lagged behind MSCI, as evidenced by their lower Sharpe and Sortino ratios. CAPM regression analysis yields beta values of 1.35 for LPX50 and 1.23 for LPXBO, indicating their respective sensitivities to market movements. The LPE indices exhibited negative but statistically insignificant alphas. The regression results also show R-squared values of 81% for LPX50 and 79% for LPXBO, suggesting a stronger correlation with MSCI in the last decade compared to the broader 30-year period.

Figure 3: Price-to-book ratios (PD indicators) versus the CISS

Price-to-book ratios for LPX50 and LPXBO against the (scaled) CISS from 12/31/2002 to 12/29/2023. The scaling factor for the CISS is 100.

3.2 Macro-financial stress, price-to-book ratios, and returns

Our visual inspection of the time series presented in Figure 2 clearly revealed the impact of various economic crises on the financial returns of the two LPE indices. Policymakers also call such time intervals periods of macro-financial stress. These periods are characterized by economic uncertainty, market volatility, and increased financial risk, affecting the broader economy and financial markets at large. Conventional wisdom claims that in periods of macro-financial stress, investor risk aversion tends to rise, leading to a decreased appetite for riskier assets. As many investors regard PE investments as riskier than more conventional assets, a shift in aggregate risk aversion can precipitate a decline in LPE share prices and, consequently, reduce returns for investors in these entities. Moreover, the portfolio companies within LPE firms' holdings may encounter financial hurdles during such economic downturns, which could further impact their performance, and, by extension, the returns delivered by LPE firms.

But it is not only LPE firms' share prices that suffer during periods of macro-financial stress. The book values of LPE firms may also be affected. If the portfolio companies experience deteriorating financial performance or if there are downward adjustments in their valuations due to adverse market conditions, it can lead to reductions in the book value of LPE firms. Moreover, impairments or write-downs may become more common during such periods, further impacting book values.

In the next step of our analysis, we examine the effects of macro-financial stress on the two LPE indices. Figure 3 depicts price-to-book ratios (the PD indicator) for LPX50 and LPXBO from December 2002 until December 2023. In addition, the figure shows the time series for the CISS financial stability index for the same period.

Table 2: Correlations between the CISS and PD

	PD AND CISS	$\Delta 1M$ (PD AND CISS)	$\Delta 3M$	$\Delta 6M$	$\Delta 12M$
LPX50	-0.786	-0.488	-0.719	-0.793	-0.826
LPXBO	-0.769	-0.493	-0.714	-0.788	-0.815

The reported figures are calculated from 253 monthly values of the price-to-book ratios for LPX50 and LPXBO, respectively, and the CISS index from 12/31/2002 to 12/29/2023. The first column reports correlations between the levels of the PD indicators and the CISS. The next four columns report correlations between absolute changes of the PD indicators and absolute changes of the CISS during the same time window. For example, the rightmost column depicts the correlation between the 12-month (absolute) change of the CISS and the contemporaneous 12-month (absolute) change of the price-to-book ratios of the two LPE indices.

In the years leading up to the onset of the GFC in 2007, the price-to-book ratio indicated an overvaluation of LPE firms, with their market values on average surpassing their book values. However, during the crisis and its immediate aftermath the market values of these firms dropped to less than half of their book values, signaling a significant undervaluation. Since then, both indices have consistently indicated that LPE firms are undervalued, suggesting that their market prices remain below what their balance sheets would imply.

A possible explanation for this persistent undervaluation since the GFC could be investor skepticism regarding the accuracy of book valuations. This skepticism might stem from concerns over the reliability of the valuations assigned to the illiquid assets held by LPE firms, which are often difficult to price accurately. As a result, investors may demand a discount to compensate for the perceived risk associated with potential overestimations of asset values on the firms' balance sheets. This discount, reflected in lower market prices relative to book values, serves as a buffer against the uncertainty surrounding the true worth of these firms' underlying investments.

The time series graphs in Figure 3 reveal that skepticism regarding the book valuations of LPE firms, leading to a demand for market discounts, intensifies during periods characterized by macro-financial stress. A visual examination of the graphs suggests a robust negative correlation between the CISS index and the PD (price-to-book) indicators, signifying that as financial stress increases, the discrepancy between market and book valuations widens. Reinforcing this observation, Table 2 presents a compilation of historical correlations between the CISS financial stability index and the price-to-book ratios, further illustrating the inverse relationship between macro-financial stress levels and the PD indicators.

The CISS index levels and the price-to-book ratios of LPX50 and LPXBO exhibit significant negative correlations, with coefficients of -0.786 and -0.769, respectively. Furthermore, the absolute changes in the CISS index and the price-to-book ratios for both indices are also strongly and negatively correlated. This implies that rises in the CISS index, signaling heightened macro-financial stress, are typically associated with reductions in the price-to-book ratios of both indices, and vice versa. Such a pattern underscores a direct inverse relationship between macro-financial stress levels and the valuation metrics of these LPE indices.

Table 3: Correlations between the CISS and the indices' returns

	RETURNS AND CISS	RETURNS AND $\Delta 1M$ CISS
LPX50	-0.242	-0.431
LPXBO	-0.239	-0.442
MSCI	-0.171	-0.342

The reported figures are calculated from 252 monthly values of the returns for the three indices and from the CISS index from 12/31/2002 to 12/29/2023. The first column reports correlations between the returns and the CISS level. The second column reports correlations between the returns and the absolute changes of the CISS during the same time window.

Table 3 shows that this inverse relationship also holds between the CISS index and the returns of both LPE indices as well as those of MSCI. Notably, variations in the CISS index exhibit a stronger (more negative) correlation with the returns of these indices than do the absolute levels of the CISS itself. This finding suggests that fluctuations in macro-financial stress, as captured by changes in the CISS, are more closely linked to the performance of the LPE indices and MSCI than the actual level of the CISS is – highlighting the dynamic impact of financial stability on market returns.

Table 4 reports correlations between CISS changes and compound index returns for three, six, and twelve months. The correlations are stronger than for the one-month time window in the right column of Table 3.

Table 4: Correlations between CISS changes and compound index returns

	3M RETURNS AND $\Delta 3M$ CISS	$\Delta 6M$	$\Delta 12M$
LPX50	-0.616	-0.673	-0.653
LPXBO	-0.607	-0.677	-0.643
MSCI	-0.553	-0.614	-0.611

The reported figures are calculated from 252 monthly values of the returns for the three indices and the CISS index from 12/31/2002 to 12/29/2023. The first column reports correlations between the three-month (absolute) changes of the CISS and the three-month compound returns of the stock indices. The next two columns report correlations for six-month and 12-month time windows.

Table 5: Correlations between lagged CISS one-month changes and compound returns

	3M RETURNS	6M RETURNS	12M RETURNS	24M RETURNS
LPX50	-0.160	-0.148	-0.119	-0.062
LPXBO	-0.182	-0.159	-0.111	-0.072
MSCI	-0.100	-0.114	-0.078	-0.030

The table reports the correlations between the absolute change of the CISS index in a month and the compound returns in the following 3, 6, 12, and 24 months for the three indices.

While the correlations between contemporaneous values of the CISS index and LPE index returns present intriguing insights into the interaction between macro-financial stress and market performance, their practical utility for trading remains limited. The simultaneous observation of these variables offers little in the way of actionable advice for forecasting future market movements. Naturally, the results prompt a critical question: can the CISS index be used not only as a coincident but also as a predictive metric that can inform investment decisions ahead of market shifts? We attempt to answer this question in the final step of our analysis.

3.3 Return predictability?

We analyze whether the CISS index could serve as a leading indicator of LPE market returns. For this purpose, we analyze correlations between lagged CISS changes and the index returns. Tables 5 and 6 report correlations between absolute changes in the CISS index and the later compound returns of the LPX50, LPXBO, and MSCI, respectively.

The correlations between monthly variations in the CISS index and the subsequent monthly returns of the three indices, as presented in Table 5's first column, align with the contemporaneous values outlined in the right column of Table 3. Changes in macro-financial stress levels are negatively correlated with the returns of all three indices in the following three months. This relationship fades over extended periods – 12 and 24 months – progressively nearing zero. This pattern indicates that the influence of macro-financial stress on compound index returns diminishes over time. Furthermore, the correlations documented in Table 6, between six-month lagged fluctuations in the CISS index and subsequent three-months returns, exhibit a comparable behavior as those observed in the first column of Table 5. For extended periods, they exhibit a similar diminishing trend. Interestingly, the correlation for the two-year compound returns of the LPE indices shows a reversal in sign, becoming positive (but is statistically insignificant). While the first three columns of the bottom half of Table 6 (correlations between lagged CISS 12-month changes and compound returns) show a similar pattern to those in the top half (correlations between

Table 6: Correlations between lagged CISS six-month changes and compound returns

	3M RETURNS	6M RETURNS	12M RETURNS	24M RETURNS
LPX50	-0.205	-0.148	-0.138	0.069
LPXBO	-0.202	-0.103	-0.075	0.107
MSCI	-0.202	-0.173	-0.165	0.003
CORRELATIONS BETWEEN LAGGED CISS 12-MONTH CHANGES AND COMPOUND RETURNS				
LPX50	-0.210	-0.183	-0.088	0.111
LPXBO	-0.175	-0.118	-0.048	0.156
MSCI	-0.186	-0.203	-0.132	-0.017

The table reports the correlations between the absolute change of the CISS index during 6/12 months and the compound returns in the subsequent 3, 6, 12, and 24 months for the three indices.

lagged CISS six-month changes and compound returns), the rightmost column shows a further reversal of the correlations for the two LPE indices.

To determine whether the observed reversal constitutes mere statistical noise, we adjust the time lag between changes in the CISS index and the compound returns of the indices. Previously, the analysis for Tables 5 and 6 used a one-month lag. We now extend this to consider a 12-month lag. For example, we examine the CISS index's change over a three-month period and relate it to the annual return of an index during the second year. Put differently, we are correlating fluctuations in the CISS index from a three-month period with the compound returns of the second year following these fluctuations. Table 7 reports the results for three-, six-, and 12-month CISS changes to the compound returns of LPX50, LPXBO, and MSCI, respectively.

While the annual return of MSCI in year 2 appears to be only weakly correlated to CISS changes over 3, 6, or 12 months, this is not true for LPX50 and LPXBO. Both indices demonstrate a positive correlation, statistically significant, between macro-financial stress over periods of six or 12 months and the annual return in the subsequent second year. These results suggest that larger macro-financial stress leads to larger annual compound returns in the second year following these fluctuations.

The analysis presented in this section offers insights into the finding from the opinion poll of institutional investors cited in this article's introduction, where 72% of respondents indicated plans to increase their allocation to private markets over the next five years. Following the macro-financial stress induced by the onset of the COVID-19 pandemic and the Russia–Ukraine conflict, investors might anticipate a rebound in LPE price-to-book ratios and robust positive returns – until the advent of the next economic downturn. Ideally, we would bolster

these indicative claims with an event study to provide more compelling evidence that low price-to-book ratios following periods of macro-financial stress are precursors to significant outperformance by LPE indices. (Un)fortunately, our dataset lacks sufficient crisis periods to permit a thorough analysis.

4. CONCLUSION

We have analyzed 30 years of return data from two well-known LPE indices, LPX50 and LPXBO. Over the entire time span, the two indices generated higher average returns than MSCI, in line with their higher volatility. Yet in the last decade, this global equity index surpassed the LPE indices in terms of risk-adjusted performance. Our investigation has also shown that post-the Great Financial Crisis LPE companies have, on average, been valued at a discount relative to their book values. This discount exhibits a strong negative correlation with the ECB's CISS indicator of macro-financial stress. In addition, the returns of the LPE indices are negatively correlated with the CISS. By employing the CISS as a predictive tool, our findings highlight that short-term fluctuations in the CISS negatively impact LPE returns in the near term. However, with a one-year lag, an uptick in the CISS metric interestingly seems to forecast a rebound in LPE performance, suggesting a complex interplay between macro-financial stress and the cyclical nature of LPE market reactions.

Table 7: Correlations between lagged CISS changes and annual returns in the second year

	$\Delta 3M$ CISS	$\Delta 6M$ CISS	$\Delta 12M$ CISS
LPX50	0.049	0.212	0.264
LPXBO	0.048	0.224	0.302
MSCI	0.024	0.160	0.128

The table displays the correlations between the absolute changes in the CISS index over periods of 3, 6, or 12 months and the annual returns of the three indices in the second year following those changes.

HIGHER CAPITAL REQUIREMENTS ON BANKS: ARE THEY WORTH IT?

JOSEF SCHROTH | Research Advisor, Financial Stability Department, Bank of Canada¹

ABSTRACT

Following the 2007-09 global financial crisis, policymakers and standard setters made an important change in how they think about the regulation of banks. While they have always been focusing on the health of banks, they now explicitly do so to make sure that there are no sudden contractions in credit supply. Consequently, success of regulatory policy is now measured not only by market liquidity or whether there are losses to deposit insurance agencies, but also by whether the supply of credit is sufficiently stable. Higher capital (buffer) requirements, paired with regulatory stress tests, are key policy innovations to support stable credit supply. These policy innovations impose costs on banks today but their intended future benefits are not well understood. This article discusses design features that determine whether the innovations' intended benefits would materialize.

1. A NEW APPROACH TO BANK CAPITAL REGULATION

Policymakers are still in the process of implementing, or phasing in, new bank regulations based on the so-called Basel III guidelines. There are two key innovations. First, regulatory capital requirements on banks are now higher on average. Second, stress tests help to determine how high capital requirements should be. Stress tests are sophisticated exercises that use granular bank level data to examine how banks would be affected in hypothetical adverse macroeconomic scenarios. They give a good idea of how banks' capital or lending would be affected in case of severe adverse economic outcomes.

Stress tests can help inform the appropriate level of additional capital (buffer) requirements levied on all banks broadly, such as in Canada, or on individual banks such as in the U.S. The idea is that when banks hold additional capital that can absorb losses during adverse times, then they should be able to maintain their lending activity better. Let us unpack this.

Conventional capital requirements force banks to reduce the size of their balance sheets when losses reduce their capital. This would be bad for economic activity, such as business investment, that relies on credit. Consequently, if banks hold additional capital buffers, on top of conventional capital requirements, then they can use those buffers to absorb losses and would not be forced to reduce lending.

So far, so good. One problem that banks face when they use their capital buffer to absorb, or provision for, losses is that their capital is now below the sum of conventional capital requirement and capital buffer requirement. In this case, capital buffer requirements typically stipulate restrictions of payouts to shareholders. But banks' primary objective is not to maintain a stable supply of loans, but to maintain a stable flow of payouts to their shareholders. It is, therefore, conceivable that banks' response to losses is not to let their capital ratio fall below the sum of conventional capital requirement and capital buffer requirement – but rather to lower their assets, which means reducing loans.

¹ Any views expressed are my own and not necessarily those of the Bank of Canada.

In other words, the main effect of Basel III reforms may be to further strengthen existing microprudential regulation, which concerns the health of banks' balance sheets, but may end up falling short of their macroprudential objective, which concerns the stable supply of loans to the economy. What a bank regulator can do to make capital buffer requirements more effective is to lower them when a severe adverse scenario, such as the ones envisioned in stress tests, materializes. As a result, banks would have to be less concerned about how maintaining lending would affect their ability to make payouts to shareholders.

When banks face uncertainty about their ability to make payouts, it reduces their shareholder value, increases their funding cost and, ultimately, lowers their ability to make loans. Regulators should alleviate this uncertainty by clearly answering two questions. First, is there a highest possible level of capital buffer requirements?² Second, what are the criteria for a reduction of capital buffer requirements? In other words, regulators need to tell banks what the "upper bound" on capital buffer requirements is, "when" requirements would be reduced, and "by how much" and "for how long". If regulators fail to communicate clearly in this way, then capital buffer requirements will needlessly create uncertainty about banks' payouts. Buffers would then be a source of dismay for both banks and regulators rather than a powerful new regulatory tool.

A straightforward way of coming up with an upper bound on capital buffer requirements is to set it equal to the hypothetical drop in banks' average capital ratio in a stress test with a particularly adverse scenario. The key is to stick with this upper bound for a substantial period of time and to not change it every time a new potential risk emerges. In particular, emerging risks related to, for example, pandemics, wars, or overall indebtedness may affect how regulators set the buffer requirement within a given range but should not affect the upper bound of that range. This is consistent with the idea that the size of buffer requirements is not the only determinant of their effectiveness: how long they are reduced also matters in terms of stabilizing banks' loan supply. Intuitively, it would be inefficient to require banks to be able to absorb losses from every imaginable risk. The cost of carrying all that capital would simply be too high for bank shareholders. In case things turn out much worse than reasonably anticipated, then the regulator can keep buffer requirements reduced for longer.

“

Banks' primary objective is not to maintain a stable supply of loans, but to maintain a stable flow of payouts to their shareholders.

”

Determining when to reduce buffer requirements is also relatively straightforward: they should be reduced when households and firms struggle to obtain loans. While there are potentially many financial indicators that can be used to measure "financial stress", a useful criterion has been formulated, in a different context, by U.S. Supreme Court Justice Potter Stewart as "I know it when I see it." For example, when bank stock prices are suddenly down and credit spreads up, then the economy is most likely experiencing financial stress.

It is comparatively more challenging for regulators to determine by how much or for how long to reduce buffer requirements. In particular, regulators will likely face the dilemma of an increase in risk at the same time as financial stress materializes. But it would make no sense to first reduce the buffer requirement, because of financial stress, and then to increase it back up, perhaps to an even higher level than before, because of heightened risks. This would only confuse banks, and financial markets, and have no beneficial effect on loan supply.

One way to address this dilemma is the following: keep the buffer requirement equal to the upper bound as long as there is no financial stress and reduce it to zero for a meaningful period of time when stress materializes. Once loan supply has recovered, regulators should require banks to rebuild capital buffers at a pace consistent with not triggering financial stress. This simple approach recognizes that it is too late to build capital buffers for risks at the time when those risks can be reliably detected by regulators, banks, or financial markets. Detecting risks associated with their balance sheets is at the core of banks' business models; hence, it is not obvious that regulators should attempt to do it for them.

² While there are many different capital buffer requirements in practice, this article refers to their sum. In fact, Sam Woods, Deputy Governor for Prudential Regulation at the Bank of England and Chief Executive Officer of the Prudential Regulation Authority, has discussed in a recent speech how the various capital buffer requirements resemble a single capital buffer requirement (<http://tinyurl.com/ycxx48hc>).

It is true that the rationale for bank regulation is that regulators evaluate risk differently from banks. Regulators, in contrast to banks, care about the broad social and economic implications of risks faced by banks, such as business failures or unemployment. As a result, they prefer banks to hold more capital for given risks. But this does not mean that regulators are better at measuring or detecting those risks. For example, when banks detect risks, they provision for expected loan losses. This reduces their capital and requires them to retain earnings to meet capital (buffer) requirements. Increasing regulatory capital buffer requirements at that point would be too late. Ideally, regulators would like banks to retain earnings before they start provisioning. But this would mean that regulators would have to be able to detect risks earlier than banks – and it is not clear how regulators would achieve this.

While banks have scope to use discretion in applying accounting rules, this does not necessarily imply a role for capital regulation. For example, following rapid interest rate increases in 2022, Silicon Valley Bank abused hold-to-maturity classification related to their bond holdings to avoid timely recognition of expected losses. In doing so, the bank had ignored that its liquidity risk in fact called into question the appropriateness of such classification choices. But, of course, at that point it would have been of little prudential benefit to raise the bank's capital buffer requirement.

The remainder of this article discusses the design and operation of capital buffer requirements in more detail. It also discusses caveats related to the credibility of the financial regulator and the impact of bank regulation on inequality.

2. OPTIMAL DESIGN OF CAPITAL BUFFER REQUIREMENTS

Choosing the size of regulatory capital buffers involves an efficiency-stability tradeoff. On the one hand, there is the efficiency loss from higher bank capital during normal times, when there is no financial stress. The reason is that banks consider capital costly and will increase loan interest rates when they are required to fund a larger fraction of lending with capital rather than with, for example, deposits. On the other hand, there is a financial stability benefit in terms of a lower frequency and magnitude of financial crises.

If we are talking about a conventional, pre-Basel III, capital requirement, then efficiency losses and stability benefits can simply be traded off against each other by calculating them separately for different levels of capital requirement. But this approach is not feasible in the case of capital buffer requirements. The reason is that the latter are dynamic in a way that responds to non-linear macro-financial linkages.



Such linkages are non-linear because the lower bank capital is throughout the economy, the stronger will a given capital buffer requirement constrain lending to the economy. Moreover, the expected path of capital buffer requirements affects banks' lending decisions today, analogous to expectations regarding future monetary policy rates.

To capture the efficiency-stability tradeoff related to capital buffer requirements, one needs to model jointly the banking sector, the bank regulator, and the overall economy consisting of firms and households. While this can be done in a relatively parsimonious model framework, capturing three elements is key. First, the banking sector makes capital and lending plans conditional on the state of the economy and on bank regulation. Second, firms rely in part on banks to fund their investments while banks rely in part on uninsured deposit funding. The latter provides market monitoring of banks whereby a bank's funding availability is positively related to its shareholder value. Funding availability has a crucial interaction with capital regulation because shareholder value not only depends on banks' capital but also on the timing of capital payouts to shareholders. Third, the regulator sets capital buffer requirements conditional on the state of the economy and on banks' capital and lending plans. It is natural to assume that the objective of a bank is to maximize its shareholder value and the objective of the regulator is to maximize some welfare criterion (such as the net present value of gross domestic product).

The model should match quantitatively important financial-cycle statistics such as the frequency of financial stress and banks' average target leverage. The former statistic can be obtained from historical (panel) data and the latter from banks' financial and regulatory reports. Stress tests can be used to gauge the size of shocks that can affect the banking sector at a given time.

Overall, the model would imply a capital buffer requirement during times when there is no financial stress as well as paths to rebuild capital buffers following a reduction of the capital buffer requirement during financial stress. Critically, the optimal paths depend on the severity of the financial stress that precedes them. Bank regulators should give banks more time to rebuild capital buffers, the more severe financial stress has been.

3. OPERATIONALIZING CAPITAL BUFFER REQUIREMENTS

The model framework discussed above produces a capital buffer requirement for given credit spreads and for given aggregate bank capital and shareholder value. But no regulator in their right mind would expect implementation to be easy. The reason is that economic models achieve internal consistency – needed to compute optimal capital buffer requirements – by making very specific assumptions about how communication takes place and about how expectations are formed. In reality, the intentions of regulators are often less clear than in stylized models. It is, therefore, necessary to carefully consider the market impact of announcing capital buffer requirements.

Any reduction of the capital buffer requirement needs to be accompanied by clear communication regarding the path of capital buffer requirements going forward. A model can help to communicate such “forward guidance”. As in the case of monetary policy, it is important to convey conditionality because the future is not known at the time that the forward guidance is given. For example, severe financial stress might be followed by capital buffer requirements that are “low for long”, which implies future capital buffer requirements that are low relative to banks' earnings. At the same time, it should be made clear that buffer requirements will be “low for longer” in case financial stress worsens.

In communicating with the banking sector, and financial markets more broadly, a bank regulator would likely adopt some of the lessons learned from monetary policy authorities. Specifically, during financial stress, a bank regulator would want to carefully calibrate its language to target a specific credit gap for given health of the banking sector (as measured by aggregate capital and shareholder value of the banking sector). If the credit gap is too large, then language about capital buffer forward guidance can be adjusted to be more accommodative, and vice versa.

3.1 Caveat: Credibility of the bank regulator

Banks' expectations about how long capital buffer requirements remain reduced following financial stress are key for the ability of a reduction in buffer requirement to alleviate financial stress. The reason is that banks consider capital to be costly. Consequently, it is necessary that bank regulators are seen as credible when giving forward guidance about buffer requirements. Banks' lending would not respond much to any reduction in capital buffer requirements that banks expect to be short lived.



It is reasonable that regulators may not wish to reduce capital buffer requirements too much during severe financial stress. For example, the capital conservation buffer is part of the regulatory capital buffer requirement stack in most jurisdictions and cannot be reduced. It imposes automatic payout (dividends and share buybacks) restrictions on banks in times of severe financial stress. The idea is that payout restrictions are very beneficial for banks' health at times when bank capital is low; their negative effects on banks' lending can be offset by promising banks capital buffer requirements that are reduced for longer.

However, initial payout restrictions may rebuild banks' capital to the point where promising banks capital buffer requirements that are reduced for longer is not necessary anymore to induce banks to lend. Banks will then have enough capital so that they provide lending that is close to socially optimal. At that point, it would be reasonable for the regulator to increase capital buffer requirements at a faster pace – to guard against future financial stress. But then the initial promise of reduced capital buffer requirements is not credible, and thus ineffective.

Banks have reasons to worry about tough payout restrictions during severe financial stress – because such restrictions make regulators' promises of reduced capital buffer requirements less credible. Regulators can address this

credibility challenge by reducing the size of constant capital buffer requirements (such as the capital conservation buffer) and instead increasing the upper bound on time-varying capital buffers (such as the countercyclical capital buffer). It would then be possible to support bank lending more during times of severe financial stress.

Regulators may impose payout restrictions during times of moderate financial stress when they do not need to reduce any capital buffer requirements (and when regulators also make no promises about doing so in the future). In such cases, there is no credibility challenge. For example, during the COVID-19 pandemic, against the backdrop of unprecedented fiscal support for much of economic activity, most major jurisdictions imposed restrictions on banks' dividends and share buybacks.

3.2 Caveat: Impact of bank regulation on bailouts and inequality

When banks have more capital ex-ante, then it is less likely that they need to be bailed out ex-post. However, it is not possible to rule out financial crises and the need for ex-post resolution and bailouts. The reason is that even though some households may be much more affected by financial crises than others, it would be prohibitively costly, in terms of social welfare, to require banks to hedge all their risk taking (just

as no household would purchase full insurance against all the risks it faces). At the same time, it is possible to consider how accounting for household inequality would affect the efficiency-stability tradeoff.

Bailouts of banks typically involve equity injections funded by the treasury department that are being repaid by banks over time. The Bagehot principle stipulates that the interest rate implied by initial equity injection and subsequent repayments should be steep, such as in the case of the Troubled Asset Relief Program during the 2007-08 financial crisis. When households differ in the amount of wealth they hold, then they may be affected differently by bailouts.

On the one hand, equity injections enable banks to maintain lending. This stabilizes labor demand of firms and the supply of deposits (that banks use to fund lending). Consequently, wages and the return on savings are stabilized in the short-run, which benefits both poor and wealthy households. On the other hand, when banks need to repay equity injections, they pass the cost of the implied steep interest rate on to borrowers. This increases the borrowing costs of firms who respond by somewhat lowering labor demand. In the long run,

therefore, wages are depressed, which especially affects poor households because they rely primarily on labor income. On net, wealthy households benefit from bank bailouts while poor households may be somewhat worse off. Taking into account adverse ex-post distributional implications from banking sector bailouts means that capital buffer requirements should be higher ex-ante.

4. CONCLUSION

Policymakers have developed a new regulatory tool designed to better insulate economic activity from fluctuations within the financial sector. The key benefit of capital buffer requirements is that they aim to constrain bank payouts rather than bank lending. However, regulators' intentions are not necessarily reflected in banks' actions. Banks may be less willing to lend when their payouts are being restricted. For the new regulatory tool to work as intended, it is crucial to take into account how banks react to it. Banks also need to know what they are supposed to be reacting to. Consequently, it is crucial that regulators have a coherent framework when communicating the timing of any payout restrictions to banks and financial markets.

FROM PATTERN RECOGNITION TO DECISION-MAKING FRAMEWORKS: MENTAL MODELS AS A GAME-CHANGER FOR PREVENTING FRAUD

LAMIA IRFAN | Applied Research Lead, Innovation Design Labs, Capco

ABSTRACT

In the ever-evolving fraud prevention landscape, mental models are emerging as a game-changer. By serving as cognitive frameworks that guide our understanding and interpretation of fraudulent activities, mental models enable financial institutions to elevate their fraud prevention efforts. This article explores the role of mental models in preventing fraud, from pattern recognition to decision-making frameworks, highlighting their significance in safeguarding assets and enhancing efficacy.

1. INTRODUCTION

The global cost of fraud, encompassing losses, prevention measures, and staffing expenses, has been estimated to be around U.S.\$5.4 trillion [Cox (2023)]. This poses a substantial hurdle for financial institutions, given that they must navigate regulatory mandates, ethical dilemmas related to artificial intelligence (AI), and intricate legacy systems. To combat fraud effectively, understanding the underlying mechanisms and motivations driving fraudulent activities is essential. In recent years, mental models have gained traction as powerful tools for enhancing fraud prevention strategies. By leveraging cognitive frameworks, financial institutions can gain deeper insights into fraud patterns, enabling proactive detection and mitigation.

2. DEFINITION OF FRAUD

In its simplest form, fraud involves a falsehood and a financial gain. Financial fraud encompasses a wide range of deceptive practices aimed at gaining an unfair advantage or financial benefit. According to Reurink (2018), financial fraud involves

the deliberate or reckless dissemination of false, incomplete, or manipulative information related to financial goods, services, or investment opportunities, violating legal stipulations.

2.1 Understanding the complex world of fraud

The insurance industry in the U.K. serves as a microcosm of the challenges posed by fraud, with significant financial implications. The insurance industry has long grappled with the daunting issue of fraud, a problem underscored by recent statistics. In 2022, the industry incurred £1.1 billion (approximately U.S.\$1.4 billion) in costs, mirroring figures from 2021 and 2020 [ABI (2022)]. Yet, there is a noteworthy decline of 19% in the number of detected fraud cases. Adding to the complexity, the average value per fraud has surged by 20% to £15,000, emphasizing the evolving nature of this challenge.

To compound the issue, fraud represents over 40% of all reported crimes in England and Wales [NCA (2023)]. However, underreporting remains a concern, with just 43% of fraud victims reporting their cases. Perceptions of triviality, concerns

about privacy, or independent resolution attempts, deter many from reporting. The consequences of insurance fraud ripple across the industry, escalating costs for insurers, leading to higher policyholder premiums, eroding trust, and posing a threat to the industry's stability.

Even though U.K. organizations invest an average of £600,000 (approximately U.S.\$ 750,000) annually in cybersecurity for fraud prevention, the problem persists.

2.2 Types of fraud

Fraud can take various forms, including hard fraud and soft fraud. Hard fraud involves premeditated schemes aimed at claiming payments for covered losses, often orchestrated by organized crime syndicates. Soft fraud, on the other hand, entails opportunistic behaviors such as exaggerating legitimate claims for personal gain. Frauds encompass a range of deceptive practices, including false financial disclosures, financial scams, and financial mis-selling.

- **False financial disclosures**, entailing deceptive statements about entities' financial status, sow an illusion of transparency while exacerbating information asymmetry [Black (2006)]. In the insurance industry, the prevalence of fraud, particularly arising from false disclosures during applications and claims, has intensified amidst the ongoing cost of living crisis. Between March 2022 and April 2023, opportunistic fraud cases surged by 61%, with motor insurance fraud comprising 51% of these cases and property insurance fraud 29%. The repercussions are severe, leading to higher premiums, policy voidances, claim delays or denials, increased financial burdens for honest policyholders, and legal consequences such as

criminal charges, fines, and reputational damage.

- **In financial scams**, fraudsters deceive individuals into voluntarily participating and handing over funds or sensitive information. They rely on lies and fabricated facts, inducing victims to make decisions based on false promises or threats [Pressman (1998)].
- **Fraudulent financial mis-selling** refers to manipulative marketing or selling of financial products, knowing that they are unsuitable for the consumer's needs. Unlike financial scams, mis-selling practices involve suggestive communication that creates misleading impressions [Pressman (1998)].

2.3 The fraud triangle

The fraud triangle, conceptualized by Donald Cressey in the 1950s, provides valuable insights into the psychological and situational factors driving fraudulent behavior. At its core, the fraud triangle consists of three key elements: opportunity, pressure, and rationalization (Figure 1). These elements offer a framework for understanding how individuals justify and engage in fraudulent activities.

Rationalization, one of the components of the fraud triangle, is particularly relevant to mental models. It involves the cognitive processes through which individuals justify or excuse their fraudulent behavior, often by minimizing its moral or ethical implications. Fraudsters employ various rationalizations to justify their actions, portraying themselves as victims of circumstances or viewing their behavior as necessary given perceived unfairness or inadequacy. This aspect of the fraud triangle highlights the role of cognitive frameworks or mental

Figure 1: The fraud triangle



Source: Cressey (1953)

models in shaping individuals' perceptions and decisions.

Similarly, the pressure component of the fraud triangle underscores the impact of external or internal forces that compel individuals to commit fraud. These pressures may include financial difficulties, personal crises, or unrealistic performance expectations. Mental models, which guide individuals' understanding and interpretation of the world around them, can influence how they perceive and respond to these pressures, potentially leading them to rationalize fraudulent behavior as a means of coping with or alleviating these pressures.

3. EXPLORATION OF MENTAL MODELS IN FRAUD PREVENTION

Mental models serve as cognitive frameworks that help individuals understand and navigate complex situations, playing a crucial role in fraud prevention.

Within the realm of behavioral science, a mental model is a cognitive framework or representation that individuals use to understand the world, make sense of information, and interpret

their experiences. It encompasses beliefs, assumptions, perceptions, and knowledge structures that influence how people perceive, analyze, and respond to situations. Mental models help individuals organize information, predict outcomes, and make decisions by providing a simplified and structured representation of complex phenomena or systems. To gain a comprehensive understanding of the mental models at play in fraud, it is essential to consider it at the micro, meso, and macro levels (Figure 2).

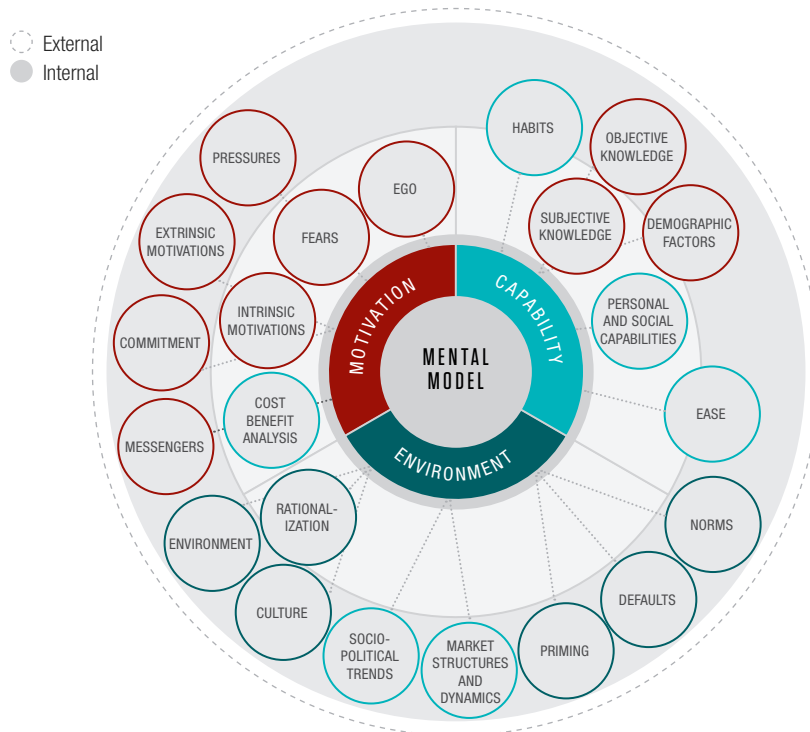
3.1 Internal factors: The micro perspective

Understanding fraud at a micro level entails analyzing demographic profiles, personality types, motivations, internal pressures, and rationalizations.

3.1.1 DEMOGRAPHIC PROFILE OF FRAUDSTERS

This involves scrutinizing age, gender, and ethnicity to gain insights into those involved in white-collar crimes. Typically, white-collar criminals are individuals over 40, predominantly male, and of white ethnicity. They often hold positions such as company owners or officers, with socio-economic standings significantly above the national average [Gekoski et al. (2022)].

Figure 2: Mental model framework



Our mental model framework looks at internal and external drivers that influence the ways in which we think and act. It outlines the complexities and interconnections between different drivers that, when understood and addressed, support development of human-centric product designs.

3.1.2 PERSONALITY CHARACTERISTICS

Fraudsters exhibit distinct traits such as overconfidence, cultural hedonism, narcissistic personality disorder, and a lack of self-control. Their behavior is often driven by a strong business mindset prioritizing competence and success, motivated by personal gain and the desire to outperform competitors. They often rationalize their actions, focusing on short-term benefits without considering larger ethical ramifications.

3.1.3 RATIONALIZATIONS

Rationalizations for fraud often include the “Robin Hood” ideology, where perpetrators justify their actions as redistributing wealth from the rich to the needy. Additionally, some individuals perceive fraud as a victimless crime, minimizing the harm inflicted on faceless entities like corporations. However, these rationalizations overlook the moral implications and real-world consequences of fraudulent behavior, perpetuating unethical actions in financial transactions.

3.1.4 OCCUPATIONAL POSITIONS

Exploring the professional roles held by fraudsters, including company owners or officers, and their socio-economic status, provides further insights into their behaviors.

3.1.5 MOTIVATIONS FOR FRAUD

Fraudsters’ motivations often exceed mere greed, as they harbor intrinsic desires for recognition and respect alongside financial gain. They may perceive their schemes as legitimate ventures akin to entrepreneurial pursuits [Frankel (2012)], fostering an illusion of trustworthiness crucial for successful investment scams [Stolowy et al. (2014)]. This mindset may also lead operators to overlook unsustainable elements within their schemes [Naylor (2007)].

3.1.6 PSYCHOLOGICAL PRESSURES

Internal pressures, such as the fear of failure and potential loss of status, are key drivers pushing individuals toward fraudulent activities. Motivations for white-collar fraudsters are complex, often fueled by a combination of factors including greed, a desire for social likability, and pressure to maintain an image of success.

3.2 External factors: The meso and macro perspective

Examining fraud from a meso and macro perspective involves analyzing affinity networks, opportunities for fraud, perceived risk and reward, control mechanisms, and socio-political and economic influences.

3.2.1 AFFINITY NETWORKS

Analysis of social, business, or personal networks used by fraudsters to identify co-conspirators and victims, fostering a sense of trust and familiarity [Stolowy et al. (2014)].

3.2.2 OPPORTUNITIES FOR FRAUD

Identification of factors such as unmonitored access to resources and unsuspecting victims that create opportunities for fraudulent activities. Access to vulnerable victims is facilitated through various means, such as “lead lists” or “mooch lists” readily available online for a nominal fee [Baker and Faulkner (2003), Nash et al. (2013)].

3.2.3 PERCEIVED RISK AND REWARD

Understanding how fraudsters perceive fraud schemes as “low-risk and high-return activities”, influencing their decision to engage in fraudulent acts.

3.2.4 CONTROL MECHANISMS

Examination of the role of law enforcement agencies, regulatory bodies, auditors, and cybersecurity experts in detecting and combating fraud. Inadequate oversight, including failures by external auditors and regulatory authorities, can enable scams to thrive [Geis (2013), Shapiro (2013), Markopolos (2010)].

3.2.5 SOCIO-POLITICAL AND ECONOMIC INFLUENCES

Analysis of socio-political trends, such as the cost-of-living crisis, housing shortages, and market dynamics, reveals potential drivers behind individuals resorting to fraud as a means of coping with financial challenges or gaining a competitive edge. Moreover, societal norms emphasizing success over ethical acquisition may incentivize individuals to prioritize financial gain regardless of the means [Trahan et al. (2005), Young (2013)]. Fraudsters adeptly exploit financial market opacity, particularly in hedge funds, leveraging technological advancements like the Internet to maintain anonymity [Frankel (2012), Blois (2013), Shapiro (2013), Stolowy et al. (2014)].

3.3 Fraudster personas

Mental models play a crucial role in shaping personas of fraudsters. Three prevalent archetypes are presented: the opportunistic fraudster, the con artist, and the trusted insider.

3.3.1 THE OPPORTUNISTIC FRAUDSTER

Opportunistic fraudsters are often seemingly law-abiding citizens. Their offenses typically involve spur-of-the-moment decisions, lacking deliberate attempts to target insurers. Instead, they stem from chance occurrences and the pressing need or desire for money. Such opportunities arise during legitimate claims processes, when introduced by others, when the fraud is relatively straightforward, or when the fraudster is emotionally unstable. The primary driver for these individuals is financial gain, often overshadowing their consideration of risk factors. Moreover, fraud is seen as an easily committed and justifiable crime, with a low risk of detection and limited police interest.

3.3.2 THE CON ARTIST

Despite their charming facade, con artists frequently demonstrate a lack of empathy, often deflecting responsibility onto their victims [Lewis (2012), Frankel (2012)]. Motivated by grandiose visions of success, they perceive their fraudulent ventures as legitimate businesses, effectively gaining the trust of unsuspecting investors [Frankel (2012), Stolowy et al. (2014)]. Scammers employ various deceptive tactics, such as exploiting trusted brand names, while preying on victims' emotional vulnerabilities, particularly greed and the desire for quick financial gains, often creating a false sense of urgency in their schemes. Genuine financial scams follow a structured approach, targeting individuals with high disposable incomes or financial vulnerabilities through promises of high returns or threats of financial consequences if they do not comply. Additionally, in romance scams, con artists utilize tactics such as mirroring, where they mimic their victims' personalities, preferences, and emotions to establish a false sense of intimacy and trust, ultimately exploiting their emotional vulnerabilities for financial gain.

3.3.3 THE TRUSTED INSIDER

Internal fraudsters often fit a profile of middle-aged white males in stable white-collar positions, as owners or officers in their companies, enjoying above-average socio-economic status [Gekoski et al. (2022)]. The link between age and white-collar crime lies in opportunity, as it takes time to

attain positions conducive to large-scale offenses. A common trait shared among offenders, regardless of their status, is salesmanship – the ability to earn trust and manipulate others, essential for both legitimate success and fraudulent activities.

Two paths diverge among offenders: one group climbs the corporate ladder through hard work and salesmanship but finds ethical compromises necessary for continued advancement, while the other achieves success but faces financial pressures that lead to fraudulent behavior to maintain their lifestyle. This latter group may feel genuine remorse for their actions once exposed, driven not by ego but fear of failure.

4. USING SITUATIONAL CRIME PREVENTION TO COMBAT FRAUD

Situational crime prevention is an approach to crime prevention that focuses on manipulating the immediate environment in which crimes occur to reduce opportunities for criminal behavior and increase the perceived risks and difficulties for potential offenders [Freilich and Newman (2017)]. This approach recognizes that crime is often opportunistic and influenced by environmental factors, such as the design of physical spaces, the presence of security measures, and the routines of potential targets (Figure 3).

Figure 3: Situational crime prevention



Situational crime prevention focuses on addressing the immediate situational factors that contribute to fraudulent activities, aiming to create environments that are less conducive to fraud and more challenging for fraudsters to exploit.

Key principles of situational crime prevention involve manipulating the immediate environment to deter criminal behavior. These tactics include:

- **Increase effort:** implementing measures that make committing fraud more difficult by adding obstacles or increasing the effort required. This could involve enhancing security measures such as multi-factor authentication, encryption protocols, or identity verification processes to create barriers and deter fraudulent activities.
- **Increase risk:** enhancing the perceived risk of engaging in fraudulent behavior by increasing the likelihood of detection, apprehension, or punishment. This might include deploying advanced fraud detection algorithms, conducting regular audits and reviews, or collaborating with law enforcement agencies to investigate and prosecute offenders, thereby making fraudsters feel more exposed and vulnerable to consequences.
- **Reduce reward:** decreasing the potential benefits or gains of engaging in fraudulent activities to make it less appealing. Strategies might include implementing stringent financial controls, conducting thorough background checks on employees and vendors, or enhancing customer verification processes to mitigate the incentives for fraudulent transactions or schemes.
- **Reduce provocations:** minimizing factors that might provoke or incentivize fraudulent behavior. This could involve implementing fraud awareness and training programs for employees and customers, enhancing internal controls and oversight mechanisms, or improving communication channels to address grievances and concerns effectively, thereby reducing the likelihood of individuals resorting to fraud as a response to perceived injustices or pressures.

By applying these principles, situational crime prevention aims to modify the immediate context in which crimes occur, making criminal behavior less attractive or feasible while promoting safer and more secure environments for individuals and communities.

Overall, situational crime prevention emphasizes proactive measures to modify the immediate environment in ways that discourage criminal behavior and promote community safety. To counteract these cognitive biases and rationalizations, situational crime prevention strategies leverage mental models to disrupt the perceived benefits of fraud and increase the perceived risks and difficulties for potential offenders.

This may involve implementing controls and safeguards within financial systems and processes to increase the effort required to commit fraud, enhancing surveillance and monitoring to increase the risks of detection, and promoting a culture of accountability and integrity to remove excuses for criminal behavior.

Furthermore, situational crime prevention in the context of fraud utilizes mental models to inform the design of intervention strategies that target specific situational factors associated with fraudulent activities. By understanding the cognitive processes and decision-making patterns of fraudsters, organizations can develop tailored prevention measures that address the underlying motivations and environmental cues that contribute to fraud.

In essence, situational crime prevention for fraud recognizes the interplay between cognitive factors, environmental conditions, and criminal behavior, and seeks to manipulate the situational context to deter fraud and promote ethical conduct. By leveraging mental models to understand the cognitive processes underlying fraud, organizations can implement targeted interventions that effectively disrupt the mechanisms driving fraudulent activities and safeguard against financial losses.

5. UTILIZING MENTAL MODELS WITHIN FINANCIAL SERVICES

Mental models facilitate pattern recognition, anomaly detection, decision making, behavioral analysis, and collaboration among financial institutions and law enforcement agencies.

- **Pattern recognition and anomaly detection:** mental models play a crucial role in financial crime prevention, particularly in pattern recognition and anomaly detection. By employing advanced algorithms and machine learning techniques, institutions can analyze financial data in real-time to identify suspicious patterns and deviations from normal behavior. This enables proactive measures to be taken to prevent fraudulent activities before they escalate.
- **Decision-making frameworks:** mental models provide decision-making frameworks that guide fraud investigators and analysts in taking appropriate action. These frameworks consider various factors, including the severity of the fraud, the likelihood of false positives, and the impact on legitimate customers. By incorporating risk-based decision-making principles, financial institutions can prioritize their resources effectively and respond to fraud incidents promptly.

- **Behavioral analysis and psychometric profiling:** in addition, mental models excel in behavioral analysis and psychometric profiling, aiding in the identification of potential perpetrators as well as vulnerable victims. By analyzing behavioral patterns and personality traits of fraudsters and victims alike, institutions can tailor their detection methods to effectively mitigate specific threats.
- **Service design:** in service design, mental models can enhance the creation of customer journeys. This involves integrating prompts and pauses to stimulate ethical reflection among customers as they navigate through the journey. By encouraging individuals to pause and contemplate the broader consequences of their actions, they can develop a deeper awareness of the ethical implications inherent in their decisions. This approach fosters a more responsible and socially conscious customer experience. Service staff can also undergo training to identify victim vulnerabilities and tactics employed by fraudsters.
- **Collaborative intelligence and information sharing:** mental models facilitate collaborative intelligence and information sharing among financial institutions and law enforcement agencies. By sharing data and insights, organizations can collectively identify emerging fraud trends, share best practices, and coordinate their efforts to combat fraud more effectively. This collaborative approach enables institutions to leverage the collective expertise and resources of the entire ecosystem, leading to more robust fraud prevention measures.
- **Continuous learning and adaptation:** finally, mental models enable continuous learning and adaptation in response to evolving fraud threats. By analyzing past incidents and identifying areas for improvement, financial institutions can refine their fraud prevention strategies and stay ahead of emerging threats. This iterative process of learning and adaptation is essential for maintaining the effectiveness of fraud prevention measures in a rapidly changing environment.

6. CONCLUSION

Most organizations have traditionally focused their fraud strategies on detection and prevention measures, often relying on technological solutions and procedural controls. However, by incorporating mental models into their approach, organizations can gain deeper insights into the motivations, behaviors, and psychological factors driving both fraudsters and victims. Understanding the cognitive biases, decision-making processes, and situational factors that influence individuals involved in fraudulent activities can help organizations develop more effective strategies for detecting, mitigating, and responding to fraud. By adopting a behavioral science perspective and leveraging mental models, organizations can enhance their fraud strategies by addressing root causes, designing targeted interventions, and fostering a culture of vigilance and resilience.

REFERENCES

- ABI, 2022, "Fraud," Association of British Insurers, <https://tinyurl.com/2bc959z7>
- Baker, W. E., and R. R. Faulkner, 2003, "Diffusion of fraud: intermediate economic crime and investor dynamics," *Criminology* 41:4, 1173–1206
- Black, W. K., 2006, "Book review: control fraud theory v. the protocols," *Crime, Law and Social Change* 45:3, 241–258
- Blois, K., 2013, "Affinity fraud and trust within financial markets," *Journal of Financial Crime* 20:2, 186–202
- Cornish, D. B., and R. V. Clarke, 2003, "Opportunities, precipitators and criminal decisions: a reply to Wortley's critique of situational crime prevention," in Smith, M. J., and D. B. Cornish (eds.), *Theory for practice in situational crime prevention, crime prevention studies*, Vol. 16, Criminal Justice Press
- Cox, M., 2023, "Fraud predictions 2024: scams, siloes and upstream polluters," Fico blog, <https://tinyurl.com/47a82nww>
- Cressey, D., 1953, *Other people's money: a study in the social psychology of embezzlement*, Free Press
- Frankel, T., 2012, *The Ponzi scheme puzzle: a history and analysis of con artists and victims*, Oxford University Press
- Freilich, J., and G. Newman, 2017, "Situational crime prevention," *Oxford Research Encyclopaedia of Criminology*, March 29, <https://tinyurl.com/2kkh85ab>
- Geis, G., 2013, "Unaccountable external auditors and their role in the economic meltdown," in Will, S., S. Handelman, and D. C. Brotherton (eds.), *How they got away with it: white collar criminals and the financial meltdown*, Columbia University Press
- Gekoski, A., J. R. Adler, and T. McSweeney, 2022, "Profiling the fraudster: findings from a rapid evidence assessment," *Global Crime* 23:4, 422–442
- IFED, 2023, "Police warn of rise in bogus insurance claims as people turn to fraud amid cost of living pressures," The City of London Police's Insurance Fraud Enforcement Department (IFED), <https://tinyurl.com/4mp4wtkv>
- Lewis, M. K., 2012, "New dogs, old tricks. why do Ponzi schemes succeed?" *Accounting Forum* 36:4, 294–309
- Markopolos, H., 2010, *No one would listen: a true financial thriller*, John Wiley & Sons
- Nash, R., M. Bouchard, and A. Malm, 2013, "Investing in people: the role of social networks in the diffusion of a large-scale fraud," *Social Networks* 35:4, 686–698
- NCA, 2023, *Fraud*, National Crime Agency, <https://tinyurl.com/jmu78m2f>
- Naylor, R. Thomas, 2007, "The Alchemy of Fraud: Investment Scams in the Precious-metals Mining Business," in: *Crime, Law and Social Change* 47, 89–120
- Pressman, S., 1998, "On financial frauds and their causes," *American Journal of Economics and Sociology* 57:4, 405–421
- Reurink, A., 2018, "Financial fraud: a literature review," *Journal of Economic Surveys* 32:5, 1292–1325
- Shapiro, D., 2013, "Generating alpha return: how Ponzi schemes lure the unwary in an unregulated market," in Will, S., S. Handelman, and D. C. Brotherton (eds.), *How they got away with it: white collar criminals and the financial meltdown*, Columbia University Press
- Stolowy, H., M. Messner, T. Jeanjean, and C. R. Baker, 2014, "The construction of a trustworthy investment opportunity: insights from the madoff fraud," *Contemporary Accounting Research* 31:2, 354–397
- Trahan, A., J. W. Marquart, and J. Mullings, 2005, "Fraud and the American dream: toward an understanding of fraud victimisation," *Deviant Behaviour* 26:6, 601–620
- Young, J., 2013, "Bernie Madoff, finance capital, and the anomic society," in Will, S., S. Handelman, and D. C. Brotherton (eds.), *How they got away with it: white collar criminals and the financial meltdown*, Columbia University Press

GLOBAL FINANCIAL ORDER AT A CROSSROADS: DO CBDCS LEAD TO BALKANIZATION OR HARMONIZATION?

CHENG-YUN (CY) TSANG | Associate Professor and Executive Group Member (Industry Partnership), Centre for Commercial Law and Regulatory Studies (CLARS), Monash University Faculty of Law (Monash Law)¹

PING-KUEI CHEN | Associate Professor, Department of Diplomacy, National Chengchi University

ABSTRACT

Central bank digital currencies (CBDCs) have gained momentum in the global financial system in recent years. Its impact on global financial regulations cannot be underestimated. Despite various motives for issuing CBDCs, the circulation of different CBDCs in the global financial networks will require central banks to revise the existing rules or formulate new ones. In this process, geopolitics has a significant role. The global financial order may head to Balkanization or harmonization. This paper discusses the counteracting forces that draw regulatory changes in opposite directions. CBDCs may change the order on payment systems, settlement and clearing mechanisms, privacy protection, capital control measures, and AML/CFT measures. Geopolitical concerns on currency sovereignty and competition over fintech innovations can simultaneously encourage cooperation and confrontation. Central banks, financial intermediaries, and the private sector should be ready to cope with significant changes in the global financial order. We argue that technological developments and geofinancial concerns will remain the predominant areas of focus for years to come. They will determine the scope and intensity of geopolitical competition, and then spill over to global finance.

1. INTRODUCTION

According to a 2022 BIS survey, 93% of central banks, representing 94% of global economic output, were engaged in central bank digital currencies (CBDCs) [Kosse and Mattei (2023)]. The Atlantic Council presents a similar picture by tracking CBDC developments in 131 countries. While there are a number of reasons why CBDCs are on the rise, their implications and potential have raised concerns regarding the direction of future global financial order.

For decades, scholars and commentators have observed the evolution of the global financial order from either geopolitical or legal-regulatory perspectives, primarily shaped by the

asymmetric power of different economies, reactions to financial crises, and even internationalism driven by various so-called global standards-setters. The implications of technological changes and innovations are, however, often underestimated. This limited dimensional lens of global policy studies in finance needs a profound rethinking. The rise of CBDC reinforces such an argument. It holds the potential to drive the global financial order to unprecedented forms of Balkanization as typically defined by the divide between West and East or North and South. Even if CBDCs lead to greater harmonization, it is nonetheless an unprecedented manifestation driven by the competition between fiat and virtual currencies, a factor ignored in the current exploration of global policy study.

¹ The authors are grateful to the excellent editorial assistance by Gabrielle Liang and Abinayan Thillainadarajah. All responsibility remains with the authors.

This paper, aiming to fill the aforementioned gap, will initially explain factors giving momentum to CBDCs. Secondly, it argues that the current developments have shown signs of a Balkanized global financial order due to the central banks' different development motives, stages of financial market maturity, regulatory attitude toward virtual or cryptocurrencies, and awareness of privacy concerns. It then highlights factors that may strengthen the harmonization of the global financial order, including inclusive CBDC cross-border experiments, common standards setting, and the seeming global consensus over a cashless society and embracement of digitalization.

The third part of the paper presents more balanced thinking by considering the shaping forces of technological change, geopolitical and financial concerns, and a significant shift in understanding the role of “soft power” in influencing the global financial order. This paper will end with policy implications and tentative response strategies.

The rise of CBDCs pose under-researched impacts on the global financial order. We argue that sovereigns explore or develop CBDCs out of various, complicated motives, including, but not limited to, the “fear of missing out”, the desire to be included in the global standards-setting process, advancing financial inclusion within their jurisdictions, or responding to the rapidly spreading fanaticism of cryptocurrencies. These motives give rise to a worldwide trend of exploring, testing, and even launching CBDCs in the minimized form of pilots. The designs, architecture, and potential cross-border circulation of CBDCs subject sovereigns to not only perceivable coordination but also anticipated divides or Balkanization.

On the other hand, the deeper sovereigns delve into the unexplored territorial waters of CBDCs, the more they realize the importance of cooperation and collaboration. If CBDCs were to cross borders and facilitate trade or money flows, then a set of bilateral, multilateral, or even universally applicable rules would be necessary. These rules may take the form of retail-wholesale settlements, multicurrency exchanges, concerted trade practices, digital wallet standards, coordinated capital in-and-out flow controls, and even widely agreed data collection and privacy protection safeguards. Yet cross-border cooperation over these issues is no simple task. Sovereigns will inevitably face difficult challenges.

2. CBDC DEVELOPMENTS SHOWING SIGNS OF BALKANIZATION

Historian and philosopher, Maria Todorova, developed the theory of Balkanism, a cultural and political reflection of its conceptual counterpart, Edward Said's “Orientalism”, and calls for a fundamental discourse about an imputed ambiguity stemming from innocent inaccuracies driven by imperfect geographical knowledge and misunderstanding about the region [Todorova (2009)]. It provides a much more nuanced reality and opposes the widely shared view that “the Balkans have been described as the ‘other’ of Europe ... [and] its inhabitants do not care to conform to the standards of behavior devised as normative by and for the civilised world.” Such a historical, political, and cultural discourse helps us to get a better understanding of the connotations associated with the term “Balkanism”. It is not equal to a somewhat biased understanding of “Balkanization”, a term that originated after the Balkan Wars of 1912-1913, which “denote[s] a process of national fragmentation of former geographical and political units into new problematic national states with disrupted political relations” [Zemon (2018)]. The term “Balkanization”, as this paper also argues, symbolises a more complex portrayal of a disadvantaged group of geopolitical inhabitants (like the Balkans) forced to react to global powers with different ways of shaping ideology and implementing strategies. Researchers, including Paul Scott Mowrer and Michel Foucher, also support this view [Longley (2022)].

In other words, Balkanization is a product of exogenous interacting dynamics between sovereigns and their agents. This understanding is paramount, as a clear recognition of the direction of certain rising global phenomena must consider ongoing external factors.

Balkanization in the global financial system is discussed in various academic literature, and perhaps the most recent and potent “financial Balkanization” account lies in Wong (2022). Wong discusses how the “Russian invasion of Ukraine and the COVID-19 pandemic have fundamentally transformed geopolitics and finance” and defines financial Balkanization as “the decoupling and recoupling of international financial ecosystems that culminates in a series of overlapping at the peripheries but separate at their core capital spheres.”

The article argues that the “wider perceptual and normative mistrust and antagonism” among sovereigns will come from trade blocks, geopolitically military confrontations, and even the uncoordinated pursuit of ESG standards and goals.

Despite the aforementioned arguments, financial Balkanization does not necessarily result in one-dimensional positive or negative consequences. For example, Coley (2015) argues that the international controversy surrounding the U.S. effort to regulate cross-border banks in the aftermath of the global financial crisis has, in fact, resulted in the need to embrace Balkanization in global finance to prevent future financial crises arising from the pursuit of single-minded international standards of banking.

Nevertheless, even in the U.S. context, we can see counterarguments that discuss how regulatory fragmentation and the Balkanization of financial markets can harm the competitiveness of the American financial services sector [Bennetts (2014)]. Bennetts argues that regulatory harmonization is necessary to prevent Balkanization and promote a more efficient and competitive financial system.

While recent commentaries present a more complex view of the preceding debate, one thing seems inevitable: today's global finance is shaped by new geopolitics [Setser (2022)].

This paper does not take a specific stand on the plausible effects of financial Balkanization; instead, it aims to identify how a rising exogenous technological development, such as CBDCs, would add additional layers of financial Balkanization. It is fair to say that CBDC is mostly an exogenous factor. This idea did not gain much traction until the then Facebook proposed its ambitious global stablecoin-like Project Libra in late 2018, which “spurred” central banks to explore CBDCs as a counterbalancing act [Duncan (2022)]. Regulators worldwide seemed concerned about the effects of big tech and new forms of payments on monetary sovereignty.

For example, Jiang and Lucero (2022) suggest that the revised version of Project Libra, Project Diem, “sped up China's experiment with e-CNY because of the perceived threat to currency sovereignty.” While central bank motivations for exploring or developing CBDCs might vary between advanced and emerging economies, the majority seem to use them with the intention to maintain financial stability, implement monetary policies, enhance efficiencies of payments systems, advance financial inclusion, and ensure payment robustness [Kosse and Mattei (2023), Laboure et al. (2021)]. However, this

paper argues that these seemingly endogenous factors are superficial and that the primary motivation, as also suggested by BIS (2022), seems to be the rise and prevalence of cryptos and stablecoins that drive a fundamental rethinking on the part of central banks about their roles, missions, and capabilities.

Interestingly, the BIS works hard to promote harmonization and interoperability between different CBDCs in cross-border flows. The BIS Innovation Hub has launched and implemented numerous projects since multi-CBDC arrangements in 2021 [Auer et al. (2021)], and such efforts continue through Project Ubin, Project Jura, Project Dunbar, Project mBridge, Project Jesper, Project Aber, Project Icebreaker, Project Mariala, Project Sellar and Project Mandela. These projects aim to promote interoperability and settlement between CBDCs in cross-border transactions. Participating jurisdictions and central banks include the: New York Federal Reserve Bank, Bank of England, Hong Kong Monetary Authority, Bank of Thailand, People's Bank of China, Central Bank of the United Arab Emirates, Saudi Arabia Central Bank, Banque de France, Monetary Authority of Singapore (MAS), Swiss National Bank, Reserve Bank of Australia (RBA), Bank of Korea (BOK), Central Bank of Malaysia (BNM), Bank of Israel, Norges Bank, Sveriges Riksbank, and South African Reserve Bank.

Despite the pace and number of these cross-border CBDC projects, their participation remains limited. This could be due to their experimental nature or perhaps some other unknown geopolitical concerns. Asian central banks dominate these projects, with the Monetary Authority of Singapore (MAS), Hong Kong Monetary Authority and the Central Bank of Malaysia (BNM) being the most frequent and active participants. In Europe, it is the Swiss National Bank that leads the way.

Interestingly, despite most commentators agreeing that China leads the world in piloting and potentially launching large-scale e-CNY, its central bank only participated in the mBridge project.

It is important to highlight that the U.S., the U.K., and Japan, the world's three most powerful financial centers, had very little participation in these efforts. It could be argued that they are simply taking a more cautious approach and experiment domestically before they are ready to transcend borders. But if that is indeed the case, then why were they so eager to create standards as early as 2020 to set out common foundational principles and core features of a CBDC? And, why was China not included in this very important standards-setting effort?

What might complicate the issue further is that there are already four monetary jurisdictions that have officially launched CBDCs, including the Bahamas, Jamaica, the Eastern Caribbean Economic and Currency Union, and Nigeria. The majority of them are located in Central America and the Caribbean area. Why did the cross-border experiments idea never occur to them? A simple explanation could be that these local initiatives are not successful and scalable, and, hence, it is too early to worry about cross-border flows. Having said that, the Caribbean Community have been working hard to integrate monetary systems and markets in the region, and when the CBDC opportunity does arrive, there is nothing to hold them back from pushing further cross-border efforts to reach integration and harmonization.

It is fair to say that the answers to the aforementioned questions remain speculative for the time being, but the genuine reasons behind these developments are never made public and global politics scholars like one of the present authors would find it difficult not to suspect that geopolitical concerns have a role to play.

The developments of the past three years have shown that CBDC initiatives are likely to be decentralized [Wang and Gao (2021)], and that currency blocs might emerge [Zhang (2020)]. It may become a fragmented CBDC bloc world, resulting in financial Balkanization.

We further argue that if such Balkanization becomes a reality, it is likely because of the following four main reasons.

First, different countries have different reasons for developing CBDCs, which could hamper harmonization efforts. For instance, motives like competing for monetary hegemony, getting rid of the “dollarization” problem, and enforcing stricter capital controls will result in sovereigns preferring to develop their own block or network, exclusive of participation from their potential competitors.

Second, not every state is at the same level of financial market maturity. Following harmonized actions and so-called “universal” standards might jeopardize a state’s financial institutions’ soundness and competitiveness, and, in the worst-case scenario, hamper the state’s financial stability. For example, after Japan was forced by the U.S. and the U.K. to follow the Basel Accord of 1988, it experienced banking turmoil, which some attributed to the Basel Accord. Despite some arguable empirical evidence refuting that accusation [Montgomery (2005)], the widely subscribed belief remains.

“
...the rise of CBDCs has shown signs of Balkanization. The only question is whether this will be counterbalanced by other factors leading to harmonization.”

Third, sovereigns are still figuring out the interacting dynamics between cryptocurrencies and fiat currencies, and may still want to harness the potential benefits of the former. Cryptocurrencies’ decentralized nature, programmability, and the power to transcend economic turmoil and forced prohibition of capital outflows during wars are making policymakers rethink their stance on crypto assets, like stablecoins. This is particularly because a war plan is no longer a remote concern, given the current geopolitical instability across the globe.

Lastly, CBDCs make it easier for central banks to fine-tune monetary policy, as they have access to granular data about countrywide transactions should they want. However, such a data collecting and analyzing practice would subject central banks to significant privacy invasion concerns [Tsang et al. (2023)]. Notably, the common understanding is that central banks almost have no interest in invading citizens’ privacy, though it is difficult to suggest that their governments have zero interest in that undertaking. For believers of surveillance capitalism [Zuboff (2017)], privacy concerns arising from CBDCs are inevitable, if not natural. Some legal constructs aimed at safeguarding citizens from privacy invasion, such as the famous General Data Protection Rules (GDPR) in the European Union, have generated “Brussel effects” and many commentators argue the extraterritorial outreaches of domestic laws would introduce regulatory fragmentations or unintended negative consequences [Gstrein and Zwitter (2021), Senz and Charlesworth (2001)]. Whether such phenomena will manifest in CBDC circulation remains to be seen, but it is foreseeable that more significant fragmented attitudes toward this issue will emerge.

CBDCs, as of the writing of this paper, remain largely experimental and not alive. It is too soon to predict whether their wider launch would necessarily subject the global financial order to a new round of fragmentation or Balkanization.

Nonetheless, the ways these projects are being developed and the concerns already raised have sowed the seeds of potential disagreements, if not confrontations. One does not need to wait until the full-fledged bloom of CBDCs to witness a Balkanized financial order. In fact, even in experimental and explorative stages, the rise of CBDCs has shown signs of Balkanization. The only question is whether this will be counterbalanced by other factors leading to harmonization.

3. CBDC AS A CATALYST TO HARMONIZATION OF THE GLOBAL FINANCIAL ORDER

As much as CBDCs may increase the risk of regulatory Balkanization, several factors may drive central banks to harmonize CBDC-related regulations. These regulations range from the technical standardization of the CBDC system to the interoperability between CBDCs. As states try to meet the challenges of currency sovereignty brought about by cryptocurrencies, their shared motive may result in macro behavior that coordinates divergent interests [Schelling (2006)]. Such coordination may require a leading state or a non-state third party. But harmonization may also start from a small group of states and gradually shape a global order as more jurisdictions follow voluntarily.

The primary impetus for regulatory harmonization is facilitating cross-border transactions. CBDCs can have various designs and can be built upon a variety of security and privacy standards based on the preferences of their home governments. If a CBDC is meant to circulate solely within one's national borders, then the central bank can simply tailor it to meet the requirements of its domestic financial markets. However, cross-border transactions involve various CBDC systems. Each currency may have its regulatory requirements. Consequently, central banks may need to balance their domestic regulatory requirements with the need to link to other CBDCs. This encourages central banks to seek a shared, commonly recognized, and coordinated, global regulatory framework.

When it comes to cross-border circulation, central banks need to establish and maintain a safe, accurate, and efficient settlement mechanism that sustains a large volume of transactions. Cross-border transactions will necessitate the regulatory requirements for CBDC security, combating counterfeit currency, and interlinking payment systems between CBDCs [Bindseil and Pantelopoulos (2022)]. On the technical side, allowing retail transactions requires an interoperable platform, compatible ID verification protocol, and equivalent privacy protection measures. Privacy standards are

more salient since different jurisdictions can have large gaps in terms of privacy requirements.

Many central banks recognize the importance of cross-border transactions. Some have conducted studies on cross-border CBDCs and completed several joint projects aimed at facilitating safe, efficient, and low-cost cross-border CBDC transactions. These projects have also tested the applicability of the new technologies used in cryptocurrencies, such as the "distributed ledger technology" (DLT). Project Jura, for instance, tested the transfer of wholesale CBDCs between the euro and Swiss franc using a single DLT platform [Project Jura (2021)]. Saudi Arabia and UAE also conducted Project Aber to investigate the management of cross-border ledger systems [(Saudi Central Bank and Central Bank of the U.A.E. (2020)].

Nevertheless, many multilateral projects have focused on the retail market due to the high volume of transactions and the requirement for system interoperability. Australia, Malaysia, Singapore, and South Africa explored the multi-CBDC settlement in Project Dunbar [Project Dunbar (2022)]. Similarly, Project Inthanon-LionRock, initiated by Hong Kong and Thailand, created a prototype platform to support multi-CBDC cross-border transactions. Phase three of that project was named Project mBridge, which aims to build a common infrastructure that settles cross-border payment with fast, secure, and low-cost settlement [Project mBridge (2022)]. The latest, Project Sela, conducted by the BIS, Hong Kong and Israel, explored a potential solution to accessibility and security risk [Project Sela (2023)].

The creation of a joint payment system, or CBDC platform, will affect regulations on CBDC settlement, ID verification systems, interbank network operation, and cyber security. In the meantime, cross-border CBDCs would result in central banks facing challenges in cyber security, settlement risk, and connections between domestic and overseas banking systems. These challenges press central banks to seek solutions. The aforementioned projects aim to solve these issues by improving interoperability. The pursuit of interoperability then stimulates regulatory harmonization.

To be sure, the various projects may suggest that some jurisdictions wish to establish a new infrastructure that inevitably competes with the existing one. This may mark the beginning of Balkanization. However, the evidence so far suggests that there is some optimism since these projects, with limited participants, do not intend to create exclusive CBDC networks. These projects favor the participating jurisdictions, but would not introduce drawbacks for non-participants. For

instance, the infrastructure being tested in mBridge aims to build a platform that applies the latest technology and may be accessible to other CBDCs. The project is open to other jurisdictions, and more jurisdictions participated in the next phase. These efforts contribute to the accessibility and security of a cross-border payment system. The mBridge Project may well become the prototype of a global CBDC infrastructure.

The issue of combating financial crime is the second driving force for regulatory harmonization. This is a top-down approach initiated by states with a clear concern about AML/CFT (anti-money laundering and combating the financing of terrorism). This would most likely have an impact on the transparency of banking supervision in global finance. CBDCs, like existing currencies, may be used in illicit activities or terrorist financing. CBDCs also have a similar propensity to cryptocurrencies, which makes tracing transactions difficult. CBDC transactions can be encrypted, are anonymous, and can be quickly made across borders. Depending on the various designs of the CBDC system, the KYC (know your customer) process of CBDC may be weak and, therefore, could potentially become a loophole in global AML/CFT efforts. AML/CFT is a problem of national security. Even countries with no, or very limited, criminal activities or terrorist threats would face pressure from other countries to build robust AML/CFT measures. This is why AML/CFT is a key concern when issuing retail CBDCs. All jurisdictions that have already issued retail CBDCs have KYC measures, or certain restrictions on commercial banks, to enforce AML/CFT [Kakebayashi et al. (2023)].

The global AML/CFT regime is administered by the Financial Action Task Force (FATF). It is a rigorous regime with strong coercive force. Unlike the typical soft law-based financial regulations, FATF has clear mandates and institutions to combat money laundering and terrorist financing. Its regional agents conduct periodic reviews on states and blacklist those that fail to comply with the AML/CFT measures. In recent years, the development of CBDCs has also caught the attention of the FATF. Given it is not clear how central banks will design CBDCs, the FATF gave advice in its 2020 report, which is quite similar to the requirements for fiat currencies [FATF (2020)].

Following the FATF's study, the global AML/CFT requirements for CBDC are likely to focus on the interoperability of different CBDC systems and the ability to trace money flows through financial intermediaries. This will affect the design choices of ID verification, privacy protection, and bridging mechanisms between CBDCs. Regulatory convergence is not FATF's main concern, but a successful AML/CFT regime will depend on global regulations that apply to every jurisdiction. For example,

the AML/CFT may require ID verification (KYC process) on cross-border transactions. It may also require commercial banks and central banks to keep transaction records. Central banks would find their hands tied when it comes to the transaction verification process, record keeping, and record sharing.

More importantly, major economies, such as the U.S. and the U.K., are likely to support the AML/CFT regime. It is in their interests to avoid CBDCs following the same path as cryptocurrencies. They also have an interest in regulating smaller economies that may host and relay illicit activities. The FATF will continue to impose the top-down AML/CFT regime on states.

The support from major states is crucial to regulatory harmonization. In addition to AML/CFT concerns, great powers also have an interest in a stable global financial system. The global financial order is largely coordinated in various intergovernmental organizations. This includes the G7, G20, BIS, IMF, and the World Bank. The regular meetings between financial ministers and central bank governors are the main source of global financial governance, where a small number of states make important decisions on financial regulations and discuss potential threats to financial stability. It is commonly recognized that the global financial order is in the hands of a few economies. States such as the U.S., the U.K., and Japan enjoy strong advantages in shaping the global financial order. Their recommendations and guidelines are specifically important to push harmonization. The U.S., in particular, has a powerful influence on the global financial network [Farrell and Newman (2019)].

As more CBDCs circulate in the global markets, the increased cyber security risk and high monetary mobility will have an impact on financial stability. The instant settlement can change the existing settlement mechanism, its competition with cryptocurrencies and stablecoins could lead to regulatory changes, and its circulation across the globe brings about national security and user privacy issues. Major economies will likely take initiatives to minimize the risks caused by CBDCs. The formation of harmonized regulations takes place in multilateral intergovernmental forums that have already started in recent years. G7 issued a set of CBDC design principles in 2021. This demonstrated the concerns of major economies regarding the development of CBDCs. The BIS Innovation Hub delivered a report to the G20 financial ministers' 2023 meeting. G20 leaders also discussed CBDCs' impact on the global economy at their summit. Similarly, IMF published an overview of its approach to CBDC capacity development [IMF

(2023)]. These examples suggest that major states are aware of the impact of CBDCs.

One can expect further measures, recommendations, or guidelines in the years to come. Although it is not clear which problems major economies would prioritize, harmonization could occur on the global settlement and payment infrastructure. Another agenda would be banking supervision requirements, which may lead to a reexamination of the Basel Accord. Regulatory harmonization most likely begins by global standard setting bodies and the AML/CFT regime, where major economies hold decision making power. This means regulatory Balkanization will raise challenges outside these standard setting bodies.

4. A NEW PERSPECTIVE: THROUGH THE LENS OF TECHNOLOGICAL CHANGE AND GEOFINANCIAL CONCERNS

It is fair to say that the current discussions have predominantly been focused on how the global financial order has been shaped and whether there is any appetite for regional fragmentations. For example, while some scholars take a harmonized global financial order as an ideal goal and argue that regional financial arrangements might pose threats to global financial governance [Henning (2017)], others question whether the economic strength of emerging powers, such as the BRICS countries, will increase the “financial multipolarity” of the current global financial order centered on the U.S. and other G7 economies [Huotari and Hanemann (2014)].

Most of the aforementioned lenses of observation have a strong focus on international politics and fall under our understanding of traditional global governance scholarship. However, CBDCs present a far more complex picture, demonstrating how geopolitics, domestic financial markets, pressure to compete with foreign counterparts, and the swift change in technology interact with one another. Among these various factors, current literature tends to underappreciate two critical ones: technology and geofinancial concerns. Lloyd and Dixon (2022) argue that a unipolar world is “dangerous to the peaceful stability of the world order and fails to appreciate the dynamic, interleaved layering across economic, trade, monetary, security, and politico-cultural functionality.” They argue that a multipolar order is needed. They further argue that “the nature and pace of technological development – driven in many cases, but not all, by the private sector – is changing the face of globalization” and highlighted that trade in nonmaterial goods is subject to rapid technological innovation via distributed ledger technology. “This digitization of trade in goods and services involves the implementation of widespread

programmable (automated) contracts. This development could be further stimulated by the future launching of CBDCs for such cross-border payment transactions.”

We would in fact go further and argue that the rise of CBDCs presents a perfect combination of the two. Turner (1943) suggested that the power of technology will result in the progress of transportation, communication, global military confrontation, or power imbalance. This line of literature analyzes how innovative weapons hold the potential to change military dynamics among states. More recent studies do not necessarily share that perspective. For example, Collins (1981) argued that “[m]odern technologies of long-distance warfare, along with modern transportation and communication, do not result in any major change in the underlying principles of geopolitics.” Despite differing views, one can hardly argue that the invention of nuclear weapons did not change how geopolitics is understood. Nuclear weapons not only concluded the Second World War, they also helped create a new form of great power competition, as well as gave rise to a set of governance structures regulating atomic energy and fissile materials. The great powers then drafted the Nuclear Non-Proliferation Treaty to consolidate their geopolitical interests, restraining other countries from challenging their position by developing nuclear weapons.

Nonetheless, in the area of global finance, how technological evolution changes interacting dynamics among countries remains a largely unexplored territory. We have, of course, seen the phenomenon of fintech positioning some countries as the leaders in functioning as global financial centers, such as the U.K. and Singapore. However, the competition among fintech centers is not significant enough to change the global financial order, unless it redirects capital flows in a drastic way. Unlike a purely innovative technological invention or application, CBDCs are deeply intertwined with global monetary territory and circulation. CBDCs also hold the potential to redirect global capital flows when they become prevalent.

The rise of CBDCs is particularly distinct in the following aspects, with mixed technological and geofinancial implications.

First, if CBDCs circulate significantly across borders, then the spillover effects must be addressed [Tsang and Chen (2022)]. One way is to instill controls of the CBDC wallets. This would require some technological design, such as specialized chips and other software safeguards. Semiconductor chips used for storing CBDCs and recording their transfer might raise national security concerns for some countries [Miller (2022)]. For instance, given the strained U.S.-China relations in recent

years, one can hardly imagine U.S. CBDCs being stored in a Huawei-designed chip and mobile phone. Such concerns would drive major economies that issue CBDCs to compete for chip technology and the standards-setting powers for technical specifications, further catalyzing change in the global order in the process. The role of technological standards in geopolitics cannot be underestimated. One vivid example is China's emergence as a major player in developing technical standards, including 5G, AI, IoT, etc., which "reintroduce[s] an element of geopolitics into what are too often considered as benign, technical processes" [Seaman (2020)].

Second, cross-border CBDCs require the countries involved to reach a consensus on clearing and settlement arrangements and enhance interoperability, be it in a retail or wholesale context. This consensus formation process would likely force the world's policymakers to rethink the need to overhaul the current correspondent banking system and even the SWIFT network. SWIFT has long been a network enabling global money transfers and serves certain policy aims. For instance, the exclusion of certain Russian banks from SWIFT ended up denying Russia access to international capital markets, which presented a major challenge to Russia's financial markets. The threat of CBDCs to SWIFT was not a remote concern in May 2021. SWIFT, in conjunction with Accenture, published a report that set out practical requirements for the adoption of digital currencies and highlighted how SWIFT can continue and extend its current role to cross-border CBDC payments [Swift and Accenture (2021)]. A potential challenge to SWIFT would surely introduce geofinancial battles among major economies. In fact, some commentators highlighted that China is exploring ways for its e-CNY to bypass SWIFT in the execution, clearing, and settlements of transactions [van der Linden and Łasak (2023)]. China also joined the mCBDC Bridge Project to explore a multicurrency cross-border payment system for wholesale activity, probably with the agenda of bringing other Asian countries on board [Sewall and Luo (2022)].

Third, CBDCs have the potential to evade sanctions [Demertzis and Lipsky (2023), Kar and Priyadarshini (2022)]. A cross-border CBDC or even a global CBDC might significantly hinder sanctioning bodies' powers as they will need to bring sanctions targeting other collaborating central banks or multinational institutions, which will cast doubts on the sanction's legitimacy when it brings negative externalities to not necessarily relevant countries or entities. Indeed, today's global financial system remains pretty much dominated by U.S. dollar primacy. However, the kind of sanctions on Russia and its ripple effects can always generate distrust toward such a dollar primacy. As Singh (2022) reminds unequivocally, "Dollar

primacy is nothing more than a network. All networks have tipping points, often psychological ones that are impossible to identify in advance." After the U.S. blocked Iran and Russia from the SWIFT network, many states, including China, raised concerns that the U.S. may cut them off from the global financial network, devastating their economies by denying them access to global financial services and markets. CBDCs become a viable alternative that does not rely on the existing network. However, sanction busting will not go unnoticed. The U.S. may find ways to secure its choke point position in global finance. The challenge and the corresponding response to the sanction regime intensify the geopolitical competition over CBDCs.

Finally, CBDCs hold the potential for the Global South to deviate from the U.S. dollar hegemony. China's e-CNY and its efforts in working closely with other countries via mBridge Project might well provide a model for the Global South [Tharappel (2023)]. Under the current global financial system, dominated by U.S. dollar primacy and Washington consensus, Global South countries are encouraged to liberalize their trade and investments with the rest of the world largely by introducing foreign investments. Yet, market reform does not necessarily guarantee development. After decades of globalization, many Global South countries suffered from trade deficits, staggered economic development, and currency devaluation. Some had to engage in dollarization to sustain their economies. However, with CBDCs, these countries can potentially control spending in more effective ways, such as programming their currencies to follow their national development priorities [Tharappel (2023)]. Such a potential challenge to the primacy of the U.S. dollar is not purely imaginary. Singh (2022) is concerned CBDCs may either "enhance or erode the potency of US economic statecraft." As the world is facing more intense geopolitical conflicts and the threats of nuclear-armed powers, resorting to military solutions might no longer be sufficient or adequate. Frequent uses of economic persuasions will become common. However, the rise of CBDCs might undermine the potential power of this long-lasting economic craft led by the U.S. Such a dynamic would call for a new global financial order.

Having analyzed the four distinct ways in which CBDCs might shape the existing financial order, we present new perspectives on observing the traditional geopolitics in international relations and global studies. We argue that a mixed consideration of technological evolution and geofinancial change is missing and urgently needed in the current discussion. CBDCs remind us of the gap and provide a perfect, though still remote, example of how the global financial order will play out in the years to come.

5. CONCLUSION: WHAT NEW GLOBAL FINANCIAL ORDER MIGHT EMERGE?

The rise of CBDCs is driven by technological developments as well as geofinancial concerns. Cross-border CBDCs have brought the global financial order to a crossroads where one path directs us to Balkanization, and the other leads the world to further harmonization. It might also well be the case that these two phenomena are taking place at the same time, shaping global financial order in an untraditional way. As a result, states are likely to compete over CBDC regulations and technological standards. Such competition may well bring about significant changes to the existing order, empowering those that monopolize the technology and governance structure and develop a robust CBDC infrastructure.

This paper identifies four main reasons favoring the Balkanization path, including the fact that CBDCs were motivated by various rationales, that states are at different stages of financial market development, that the interacting dynamics between fiat and cryptocurrencies remain unsettled by many sovereigns, and that some states might utilize CBDCs to fine-tune their monetary policy or even enable new forms of surveillance capitalism. To present a balanced view, we also highlighted that harmonization could be driven by factors such as central banks' emphasis on the importance of smooth cross-border transactions, the enforcement of AML/CFT rules, and the desire to achieve a more stable financial system.

Looking beyond traditional geopolitical factors and a global policy lens, we argue that a new perspective of observing how the global financial order might be shaped by CBDCs is needed. We argue that technological evolution and geofinancial implications added additional complexity to the current global financial order and hold the potential to reshape the order in four distinct ways: calling for a need to address spillover effects, regrouping states in the name of achieving interoperability, assisting in evading sanctions and running afoul of traditional western power of non-military actions, and enhancing the power of the Global South in competing with the monetary hegemony led by the U.S. and the Global North.

Similar to the way semiconductors and AI intervene in international politics, states are likely to take CBDCs and relevant technologies as an advantage in great power competition. The great powers will take measures to secure their advantages in CBDC development. The problem is not simply about issuing a retail or wholesale CBDC. What matters is the system a CBDC operates on and the global financial

network that supports cross-border transactions. In light of new geofinancial changes, central banks will take on more responsibilities that are not their forte. Geopolitical concerns are present in every decision related to CBDCs. Intervention from the executive branch is expected to become more frequent. It is quite different from the mode of cooperation between central bank governors and financial regulators. They will meet more challenges based on political assessments rather than economic ones. As they try to modify or build a global financial order, their engagement will be shaped by both competition and coordination. The order they attempt to create, either Balkanized or harmonized, may be more volatile than it used to be. Regulatory guidelines will change more frequently as new technology continues to emerge.

What it means to the private sector is the occurrence of more uncertainty about global regulatory standards. This is particularly true if they operate across different jurisdictions. Geopolitical competition places hurdles in the way of business opportunities, and geofinance competition affects their access to foreign markets and capital. They may also face more stringent and complex banking supervision. Capital control measures will likely tighten as well. On the other hand, the private sector is a beneficiary of CBDCs. Cross-border transactions will be more efficient and reliable. More business opportunities will be available, which may present new business models.

This paper argues that CBDC is much more than an alternative means of exchanging commodities and services or a tool to advance financial inclusiveness. Developing CBDCs is hardly a domestic matter, especially for major economies in the world. Once a major economy launches a CBDC and circulates it globally, it will soon have an implication on global financial governance. It, therefore, needs more caution and planning. This might explain why developed economies are relatively cautious about launching CBDCs and why the four jurisdictions that have launched them are all developing countries. Even China's e-CNY is still in the pilot testing phase; nevertheless, it is expected that if China eventually launches e-CNY officially, or even just at a larger scale, then other major economies will have to respond and follow suit in the near future. Their CBDCs will be a catalyst for the next global financial order. Whether the world will head to a Balkanized or harmonized order largely depends on the competition between states in global standards-setting bodies. Central banks, financial intermediaries, and private sector players should all get ready for the upcoming turmoil in the global financial order.

REFERENCES

- Atlantic Council, "Central bank digital currency tracker," <http://tinyurl.com/425h7pyz>
- Auer, R., P. Haene, and H. Holden, 2021, "Multi-CBDC arrangements and the future of cross-border payments," BIS papers, no. 115
- Bank of Canada, European Central Bank, Bank of Japan, Sveriges Riksbank, Swiss National Bank, Bank of England, Board of Governors of the Federal Reserve, and Bank for International Settlements, 2020, "Central bank digital currencies: foundational principles and core features (report no 1)," BIS Innovation Hub Other, <http://tinyurl.com/4x767czz>
- Bennetts, L., 2014, "Regulatory fragmentation, the Balkanisation of financial markets and the competitiveness of the American financial services sector," Cato Institute, <http://tinyurl.com/2atjy4dk>
- Bindseil, U., and G. Pantelopoulou, 2022, "Towards the holy grail of cross-border payments," ECB working paper no. 2022/2693
- BIS, 2022, "BIS annual economic report 2022," Bank for International Settlements
- BIS Innovation Hub, 2021, "BIS Innovation Hub work on central bank digital currency (CBDC)," <http://tinyurl.com/56zmh9z6>
- Coley, A., 2015, "U.S. regulation of cross-border banks: is it time to embrace balkanisation in global finance?" SSRN, <http://tinyurl.com/mrx65wez>
- Collins, R., 1981, "Does modern technology change the rules of geopolitics?" *Journal of Political & Military Sociology* 9, 163–177
- Demertzis, M., and J. Lipsky, 2023, "The geopolitics of central bank digital currencies," *Intereconomics* 58, 173–177
- Duncan, E., 2022, "Central banks 'spurred' by Facebook's Libra to explore CBDCs. Open Banking Expo, <http://tinyurl.com/mw5ad82n>
- Farrell, H., and A. L. Newman, 2019, "Weaponized interdependence: how global economic networks shape state coercion," *International Security* 44, 42–79
- FATF, 2020, "FATF report to G20 on so-called stablecoins," <http://tinyurl.com/5yupvdk>
- Gstrein, O.J., and A. J. Zwitter, 2021, "Extraterritorial application of the GDPR: promoting European values or power?" *Internet Policy Review* 10
- Henning, C. R., 2017, "Avoiding fragmentation of global financial governance," *Global Policy* 8, 101–106
- Huotari, M., and T. Hanemann, 2014, "Emerging powers and change in the global financial order," *Global Policy* 5, 298–310
- IMF, 2023, "IMF approach to central bank digital currency capacity development," *International Monetary Fund policy paper no. 2023/016*, <http://tinyurl.com/24t8mdtb>
- Jiang, J., and K. Lucero, 2022, "Background and implications of China's E-CNY," *University of Florida Journal of Law & Public Policy* 33, 237
- Kakebayashi, M., G. P. Presto, T. Yuyama, and S. Matsuo, 2023, "Policy design of retail central bank digital currencies: embedding AML/CFT compliance," SSRN, <http://tinyurl.com/y3u38m99>
- Kar, S., and D. Priyadarshini, 2022, "Russia-Ukraine: could central bank digital currencies help countries bypass sanctions?" *NDTV Profit*, <http://tinyurl.com/4rr59vf5>
- Kosse, A., and I. Mattei, 2023, "Making headway - Results of the 2022 BIS survey on central bank digital currencies and crypto," *BIS papers* no. 136, <http://tinyurl.com/3w9kms3>
- Laboure, M., M. H.-P. Müller, G. Heinz, S. Singh, and S. Köhling, 2021, "Cryptocurrencies and CBDC: the route ahead," *Global Policy* 12, 663–676
- Lloyd, M., and C. Dixon, 2022, "A future multipolar world," *Global Policy* 13, 818–827
- Longley, R., 2022, "What is balkanisation?" *ThoughtCo*, <http://tinyurl.com/4b52cxja>
- Miller, C., 2022, "Chip war: the fight for the world's most critical technology," *Simon and Schuster*
- Montgomery, H., 2005, "The effect of the Basel Accord on bank portfolios in Japan," *Journal of the Japanese and International Economies* 19, 24–36
- Project Dunbar: international settlements using multi-CBDCs, 2022
- Project Jura: cross-border settlement using wholesale CBDC, 2021
- Project mBridge: connecting economies through CBDC, 2022
- Project Sela: an accessible and secure retail CBDC ecosystem, 2023
- Reuters, 2020, "Facebook's renamed cryptocurrency is still 'wolf in sheep's clothing': German Finance Minister," December 7, <http://tinyurl.com/5ba2hvt5>
- Saudi Central Bank, Central Bank of the U.A.E., 2020, Project Aber Final Report
- Schelling, T. C., 2006, *Micromotives and macrobehavior*, W. W. Norton & Company
- Seaman, J., 2020, "China and the new geopolitics of technical standardization," *Notes de l'Ifri*, <http://tinyurl.com/4mecec5c>
- Senz, D., and H. Charlesworth, 2001, "Building blocks: Australia's response to foreign extraterritorial legislation," *Melbourne Journal of International Law* 2, 69–121
- Setser, B. W., 2022, "The new geopolitics of global finance," *Council on Foreign Relations*, <http://tinyurl.com/mrtw2646>
- Sewall, S., and M. Luo, 2022, "The geopolitics of digital currency," *Belfer Center for Science and International Affairs, Harvard Kennedy School*, <http://tinyurl.com/mrym9xem>
- Singh, D., 2022, "It's not just the economics: why US leadership on CBDCs is a national security imperative," *Harvard National Security Journal* 14, 49
- Swift, Accenture, 2021, "Exploring central bank digital currencies: Swift and Accenture publish joint paper," <http://tinyurl.com/5n8xkec6>
- Tharappel, J., 2023, "How Chinese fintech threatens US dollar hegemony," *Frontiers in Blockchain* 6, 1148315
- Todorova, M., 2009, *Imagining the Balkans*, Oxford University Press
- Tsang, C.-Y., Y.-P. Yang, and P.-K. Chen, 2023, "Disciplining CBDCs: achieving the balance between privacy protection and central bank independence," *Northwestern Journal of Law and Business* 43:3, 235–289
- Tsang, C.-Y., and P.-K. Chen, 2022, "Policy responses to cross-border central bank digital currencies—assessing the transborder effects of digital yuan," *Capital Markets Law Journal* 17, 237–261
- Turner, R., 1943, "Technology and geopolitics," *Military Affairs* 7, 5–15
- van der Linden, R. W. H., and P. Łasak, 2023, "The digitalization of cross-border payment systems and the introduction of the CBDC," in van der Linden, R. W. H., and P. Łasak, (eds.), *Financial interdependence, digitalization and technological rivalries: perspectives on future cooperation and integration in Sino-American financial systems*. Springer Nature Switzerland
- Wang, H., and S. Gao, 2021, "The future of the international financial system: the emerging CBDC network and its impact on regulation," SSRN, <http://tinyurl.com/dtptxj2>
- Wong, B., 2022, "The era of financial Balkanisation," *Yale Journal of International Affairs*, <http://tinyurl.com/3fkwwk97>
- Zemon, R., 2018, "Us," "them," and the problem with "Balkanisation," *Global-e Journal* 11:16, <http://tinyurl.com/4bpc4j6>
- Zhang, T., 2020, "New forms of digital money: implications for monetary and financial stability," IMF, <http://tinyurl.com/2hkptcw>
- Zuboff, S., 2017, "The age of surveillance capitalism: the fight for a human future at the new frontier of power," *Ingram Publisher Services*

ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES

CHARLES KERRIGAN | Partner, CMS

ANTONIA BAIN | Lawyer, CMS

ABSTRACT

The integration of artificial intelligence (AI) systems within the financial services industry has the potential to transform business operations, improve customer relations, and enhance regulatory compliance efforts. However, its adoption is not without risk; the integration of AI raises significant ethical concerns and threatens market integrity, data privacy, consumer protection, and other modern tenets of law. While these concerns are not necessarily new to the financial services industry, they do present barriers to the incorporation of AI technology. This article explores both the benefits and risks associated with AI in the context of financial services, discussing the relevant policy considerations and current regulatory landscape. It synthesizes current research and industry invites to provide an overview of the opportunities and challenges associated with the use of AI within financial services while addressing the lack of certainty currently observed in formulating an approach for broader incorporation. In doing so, this article offers valuable insights for financial professionals and researchers in navigating the rapidly evolving landscape of AI-driven financial services.

1. WHAT IS ARTIFICIAL INTELLIGENCE?

Artificial intelligence (AI), for present purposes, can be defined as algorithmic and/or machine-based systems with the capabilities to carry out functions that would otherwise necessitate human thinking or intervention.¹ Essentially, it represents the combination of machine-learning and robust datasets to enable software to show learning, adaptability, and perform cognitive tasks (including problem-solving and decision-making functions, among other things).²

In practice, AI can be considered in specialized sub-categories, with each allowing for different operational outcomes and purposes. For example, predictive AI adopts a statistical analysis of past patterns and events in order to predict future outcomes. Generative AI (GenAI) considers large quantities of inputted data to produce new outputs, such as recommendations or answers to inputted questions.

The increasing speed of adoption of new AI systems provides opportunities for efficiency in terms of time, cost, and outcomes; however, its adoption is not without risk. While many of these risks are not new, there is a degree of uncertainty in the application of AI across various industries; as such, its rapid and widespread integration may attach new challenges for regulators which, in turn, may create barriers to the effective implementation of the technology. The following shall consider the adoption of AI across the financial services sector, focusing on its use-cases and the regulatory landscape.

2. HOW IS AI RELEVANT TO FINANCIAL SERVICES?

The financial services sector is subject to industry-specific regulation, leading to some natural reluctance among industry participants in adopting innovative technologies; as such, the initial uptake of AI was cautious. However, AI systems perform well in tasks that are core to the activities of financial

¹ <https://tinyurl.com/2bk6s27n>

² <https://tinyurl.com/2e53h75x>

AI in financial services

AI has and will continue to observe increasing capital investments and annual growth:

- A recent survey shows that 42% of 56 U.S. financial services executives plan on increasing AI investments by at least 50%.
- AI in financial services has a predicted annual growth of 20-34% in the Middle East.
- According to KPMG, 84% of UK financial services business leaders say that AI is at least moderately to fully functional within their organization.³

Such growth is at least partly attributable to the continuing development of the technology underpinning AI, which continues to improve upon AI's understanding and generative activities. Public Alpha chatbot exemplifies the increasing sophistication and power behind AI. To expand, the model is underpinned by approximately 1.2 billion parameters, all of which support the chatbot to engage in its processing functions, generate responses, and even grasp nuance. These functions and the increasing parameters are leading to outputs that are "indistinguishable from those a human might produce."

institutions. A recent study by U.K. Finance showed that 91% of financial institutions have now deployed some level of predictive AI in fraud detection and back-office functions, with recorded benefits. To this end, financial services firms continue to embrace different forms of AI to optimize their operations and enhance customer services. For example, AI is now widely used to leverage data, automate tasks, and deliver personalized services to clients, with common applications including:

- the deployment of chatbots and robo-advisors
- fraud and money laundering detection
- know your customer (KYC) checks
- creditworthiness assessments for loans and mortgages (with examples of banks in the U.S. adopting GenAI solutions to support with small business lending)
- automation of insights from earnings transcripts and analysis of data in investment management.

Broadly speaking, such applications have the potential to significantly improve the operational outcomes for both businesses and consumers, while concurrently limiting various risks commonly associated with the financial services industry. In addition, they may serve to support the regulatory compliance efforts of financial institutions through promoting operational resilience and facilitating firms' consumer duty.

In addition to the aforementioned operational enhancements, AI is transforming the business models of financial institutions. Service providers now offer "AI as a service" (AlaaS) to financial services firms; this involves a cloud-based AI outsourcing solution that enables organizations to adopt and test AI systems without incurring significant capital expenditure and without assuming many of the risks. In turn, financial institutions are integrating AI and machine-learning solutions into their supply chain, marking a shift from traditional business-to-business (B2B) or business-to-consumer (B2C) models to more complex structures like B2B2C or B2B2B. This evolution involves financial institutions acting as intermediaries, procuring AI solutions from third parties and bundling them into comprehensive product packages for clients. This shift not only reflects the industry's commitment to technological advancement but also underscores the importance of collaborative ecosystems in the modern financial landscape.

The below sets out two key use-cases of AlaaS, demonstrating the practical efficiencies to be derived from AI integration in FS.

3. RISKS AND ETHICS

The underlying risks and ethics of AI systems have been central to discussions on their application in virtually all industries, including in financial services. The Bank of England (BoE) recently reported that the risks presented by AI in the context of financial services can be considered under three categories, namely: (i) data, (ii) models, and (iii) governance. For present purposes, these risks will be considered in terms of those that are already seen within financial services and those that may be introduced with the adoption of AI.

3.1 Traditional finance

As an innovative technology, AI presents new challenges for regulators and industry participants; however, it also adds uncertainty and may exaggerate traditional industry risks. For example, the financial services industry is inherently subject to "bad actor" risks; these include instances of

³ <https://tinyurl.com/bdftwd5x>

AI and fraud detection

AI integration has the potential to improve operational efficiency and practical outcomes as it may detect instances of fraud before they are carried through. To the extent that card and digital wallet payments are projected to account for 86% of payments by 2026,⁴ and insofar as fraud cases continue to rise, the application of AI in fraud detection will likely prove of significant utility.

To expand, the incidence of fraud in the financial services industry continues to increase in prevalence. The Identity Theft Recourse Centre found a 78% increase in data compromises between 2022 and 2023, while Deloitte found a 90% increase in P2P payment fraud losses between 2021-2022. In other words, card fraud losses are in excess of U.S.\$33 billion per year.

Various financial services firms have incorporated AI fraud detection software to varying degrees. Most of these systems rely on “synthetic minority oversampling techniques” (SMOTE), whereby synthetic examples of fraud cases (i.e., the minority of cases) are used to balance the dataset. Through focusing solely on fraud cases, the model addresses concerns observed in traditional

detection mechanisms, namely, where cases of fraud were not identified. However, this model proves to be overly responsive in its detection insofar as it is predicated on information relating to fraud cases; in practice, this has led to too many cases of potential fraud being identified with the model producing a number of false positives. Such false positives inhibit the efficiency of transactions and have resulted in annual losses of U.S.\$443 billion to merchants.

In response to the increasing incidence of fraud and faults identified in the current AI detection methods, Mastercard has released Digital Intelligence Pro. This is an in-house-built AI model that has been developed to detect fraud while minimizing the incidence of false positives and ensuring market efficiency. It utilizes a “recurrent neural network” (RNN); having received the data from approximately 125 billion transactions flowing through Mastercard, the AI is trained to detect fraud within a multitude of transaction types (rather than solely focusing on instances of fraud). In doing so, it appears to reduce the bias that has previously led to shortcomings in AI analysis, with evidence suggesting that (at its current state of development) the Digital Intelligence Pro has the capacity to improve fraud detection rates by 20%.

market manipulation, insider threats, and cybersecurity threats, among others. Introducing AI to bad actors may serve to heighten such risks; in our cybersecurity example, hackers may leverage the machine learning presented by AI to enhance the efficiency and sophistication of their attacks. Further, it can permit instances of market manipulation and insider threats insofar as datasets may be tampered with to produce outcomes benefitting specific persons.

Similarly, data and consumer protection risks persist. AI systems may interact with and process customer data to produce outcomes that adversely affect such customers; such outcomes include, but are not limited to data leaks, discrimination, and unfair treatment of consumers.

However, the aforementioned risks all existed in some form prior to the integration of AI. Further, such risks will continue to exist insofar as they are a product of the industry’s substantive operations and outcomes. In turn, existing regulation (as

applicable to traditional financial services) may prove sufficient in addressing such risks, irrespective of the added uncertainty presented by AI.

This is not to say that AI does not present its own challenges;⁵ rather, it highlights that the risks AI simply exaggerates may be sufficiently addressed through existing legal provisions.⁶ The E.U. AI Act purports to address some of these concerns in more detail, focusing on the mitigation of some of these risks; as discussed further in Section 6, the Act shall apply as overarching regulation, covering both general and industry-specific risks associated with AI systems.

3.2 The risks associated with AI

As a developing and innovative technology, AI adoption presents unique ethical considerations and risks. Relevant stakeholders have formulated various standards for ethical AI use, including transparency and accountability, along with

⁴ <https://tinyurl.com/ymb4z3bu>

⁵ See Section 3.2.

⁶ See Section 5.

other considerations;⁷ however, to date, there has been little in the form of directly applicable legal standards. Accordingly, where existing legal regimes prove insufficient, such risks will persist and may create barriers to the effective implementation of AI in practice. The below will set out some of the perceived risks associated with the adoption of AI specifically. This is a non-exhaustive list and remains subject to change as the technology develops.

3.2.1 LEGAL UNCERTAINTY AND ALLOCATION OF LIABILITY

First and foremost, there is a lack of certainty as to the bounds of control and the legal categorization of AI; this issue has been observed even in jurisdictions where we have seen text of directly applicable AI regulation. Naturally, this creates uncertainty as to the proper allocation of liability which, in turn, creates barriers in the adoption of the technology.

While it is apparent that AI has not yet been attributed separate legal personality, there remains uncertainty in practice as to the appropriate attribution of responsibility. This is largely due to the complexities associated with the technology; AI is predicated on machine learning (i.e., it removes the need for human intervention), which implies that the outcomes are, in the most direct sense, not reliant on the actions or omissions of a person. While it could be argued that human intervention has been necessary in the development of the technology, the issue remains with whether the provider or developer can be deemed to owe a duty or obligation towards the claimant. In some instances, the answer may be clear (particularly where contractual arrangements are involved); however, in others, and particularly as the technology advances, the acts or omissions may be deemed too remote for the provider or developer to be held liable.

Further, there is often a lack of transparency and opacity in the parties responsible for the underlying AI; thus, actually determining the identities of the parties potentially responsible for the harm may prove fruitless in itself.

Without any statutory or contractual rights, those who have suffered harm due to interactions with AI have limited recourse. They may seek redress through traditional routes, such as tort; however, without clearly defined obligations and allocation of responsibility, the aforementioned complexities will create barriers to proving a viable action. In this sense, practical issues have played a part in preventing legal

Regulatory technology and supervisory technology

Regulatory technology (regtech) involves the use of technology (including the aforementioned cloud-based integrations) that purport to improve the efficiency of **financial services institutions** in managing their regulatory risk and complying with their regulatory obligations. For example, such technology can support financial services firms with regulatory and audit reporting, in producing business impact assessments and continuity plans, as well as in their AML processes.

Supervisory technology (suptech) is adopted by **supervisory authorities** in managing their regulatory compliance efforts. In this context, authorities can use suptech to support their operational and administrative efforts, such as data analysis in transaction reports to regulators as are required to be provided by regulated firms. It can also facilitate in regulatory reporting (through standardization and automated validation), compliance and market monitoring, and in the determination of risk across various industries.

certainty. Any claims for damages caused by an interaction with AI systems would likely prove prohibitively expensive and time-consuming, with the likelihood of success proving too uncertain to justify such costs. Accordingly, the courts have had limited opportunities to clarify the legal position and such uncertainty persists.

This lack of certainty creates concerns for organizations in incorporating AI systems, with liability concerns being found to be the most relevant external obstacle in the corporate adoption of AI.⁸ To expand, organizations face the risk of assuming liability for claims brought due to harms caused by AI systems, which may deter them from incorporating the technology. Further, both consumers and businesses bear the risk of uncompensated harm; naturally, this will undermine trust and confidence, acting as a barrier to incorporation. From this, it is clear that a greater degree of legal certainty and improved transparency requirements will be necessary in ensuring efficient and effective practical outcomes.

⁷ <https://tinyurl.com/4u4wtsmd>; <https://tinyurl.com/yt7tjwn3>; <https://tinyurl.com/6cptdah>

⁸ EUR-Lex – 52022PC0496 – EN – EUR-Lex, Explanatory Memorandum, <https://tinyurl.com/2s3pbp6x>,

In traditional practice, the financial services industry has sought to resolve such issues through regulation. In the U.K., the financial services industry is subject to the Senior Managers Regime, industry principles (as discussed in Section 5), and various other forms of regulation. For example, the Listing Rules require companies to make certain disclosures and seek to maintain transparent, fair trading practices. While the introduction of AI systems may add opacity to the financial services industry, it is submitted that proper legislative intervention (similar to that proposed by the E.U. AI Act, as discussed in Section 6) may serve to mitigate the aforementioned confusion.

3.2.2 ROBUSTNESS AND SAFETY

(i) The underlying dataset

As noted, AI has the potential to bring significant operational efficiencies (such as fraud detection) and may support in financial services functions and outputs;⁹ however, industry participants (including the BoE) have expressed concerns that such integration could implicate the soundness of firms that choose to adopt the services. In practice, AI systems may produce inaccurate outputs. This is not unique to AI, rather the risk exists due to faulty datasets; however, the involvement of AI means that the erroneous outputs could prove to be more widespread and persistent than if they had occurred due to human error. In practice, these faulty outputs could lead to significant and even systemic harms; for example, consistently inaccurate determinations of credit risk could lead to “inaccurate capital modelling”.¹⁰

Further, many AI systems are programmed to be adaptable insofar as they are continuously learning from the inputted datasets; while this allows for flexibility in outputs, it exaggerates the risks of data and concept drifts (and, therefore, the risk of invalidating the data model). As identified by the BoE, if an AI system is found to be insufficiently transparent or too complex, then there is a high likelihood that prudential risks (including credit and operational risks, as well as systemic risks) will arise. Naturally, such risks threaten the integrity of financial services businesses and pose significant risks to consumers.

These risks may be mitigated by the Principles for Effective Risk Data Aggregation and Risk Reporting requirements (the BCBS), at least to some degree.¹¹ Essentially, the BCBS

requires financial institutions to establish and implement robust governance and oversight mechanisms designed to ensure effective data aggregation and reporting.

Financial institutions are responsible for ensuring that any AlaaS providers they engage will comply with such requirements; this is required per financial regulation outsourcing rules insofar as financial institutions must implement various

Case study: oxyML LLC

One of the primary areas of inspection of the FCA – and other regulators in many other markets – is whether a given asset allocation at a managed fund is consistent with the stated goals and risk levels discussed in their offering documentation. This can be seen in CP 19/5 and tangentially in parts of the Investment Funds Prudential Regime (IFPR) and Internal Capital Adequacy and Risk Assessment (ICARA). Increasingly, firms are being asked to provide more data and analysis to support their level of risk taking and justify allocations to different assets. This is a challenge for many firms, which have deprioritized data services to back-office compliance and documentation relative to pre-trade allocation analytics. This continues to be a challenge as firms grapple with legacy software not designed for extensive external data reporting.

When properly implemented, AI provides an opportunity to significantly enhance back-office activities by feeding in proper data and setting appropriate constraints. Proper implementation is far from straightforward, as base natural language processing systems such as ChatGPT will report factually inaccurate information that at first glance appears correct.

oxyML's Voltsail system was able to circumvent these issues combining patented constrained optimization algorithms with heavily restrictive rules-based logic systems, resulting in verifiable, zero-trust automated documentation and compliance support. As a result, oxyML was able to ensure proper management and support of billions of dollars in assets at partner asset management institutions across the U.S. and the U.K.

⁹ See Section 2.

¹⁰ Bank of England, 2022, “Artificial intelligence and machine learning,” at 3.17, <https://tinyurl.com/47xds9dh>

¹¹ BIS, 2013, “Principles for effective risk data aggregation and risk reporting,” <https://tinyurl.com/mvsvx7ej>

procedures and oversight checks before and during any engagement with a third-party service provider.¹² In practice, this acts to ensure that recorded and inputted data is likely to be accurate and, therefore, risks attributed to data faults are somewhat mitigated; however, to the extent that this cannot be guaranteed, this remains a point of concern.

The E.U. AI Act also aims to address these concerns insofar as it creates a requirement for human oversight.¹³ Briefly, AI systems will need to be developed in such a way that they can be “effectively overseen by natural persons”; in effect, this follows the policy aims of the BCBS insofar as such human oversight should reduce the risk of poor or inaccurate data.

(ii) Market stability and integrity

In principle, AI promises to promote and protect market integrity within financial services; the technology may be used to facilitate market surveillance (detecting instances of non-compliance) while concurrently allowing firms and regulators to assess and manage market risks. However, its adoption also poses a threat to such integrity. For example, bad actor risks could result in data breaches, misuse of assets, or widespread losses. Flash crashes caused by high-frequency trading algorithms (as facilitated through AI) may destabilize the financial markets and disrupt typical trading operations.

The concentration of the best AI systems within a small number of firms may threaten competition, lead to data monopolization, and create predatory, opaque pricing strategies. Naturally, this threatens the integrity of markets and creates significant risks for consumers. Additionally, any overreliance on AI systems and algorithms could amplify the manifestation of conventional systemic risks, particularly where such technology is concentrated; here, a system or technology crash could render the interconnected, interoperable markets the subject of significant losses.

Once again, these are not new risks; rather, they attach to the adoption of any technology. In the U.K., the Financial Conduct Authority (FCA) is tasked with “protect[ing] the integrity of the UK financial system”;¹⁴ as such, there is an existing infrastructure in place whereby such concerns can be overseen by a regulator. The E.U. AI Act also aims to address the manifestation of such risks (specifically systemic risks) through regulating specific AI models that have the greatest potential to attach systemic risks.¹⁵

4. POLICY: LESSONS TO BE LEARNT

4.1 Policy considerations in financial services

When considering risk management in the financial services industry, it seems prudent to reflect on the policy considerations that were developed in the aftermath of the 2007/2008 financial crisis. The crisis exposed a number of systemic risks and shortcomings within the financial services industry, with the lessons derived therefrom proving of general and continuous relevance to the industry. In practice, the legislature should bear such policy considerations in mind when regulating the integration of AI systems within financial services insofar as such integration presents similar risks to those observed prior to the crisis. Accordingly, it is submitted that the following policy considerations should be front-of-mind in the legislative process:

- **Transparency:** prior to the financial crisis, financial instruments were deemed too complex and opaque, thereby blurring the risks associated with the products. As noted, AI systems and structures are often complex and opaque, thereby limiting the ability of courts, regulators, and consumers to determine the risks attached to their use.
- **Data quality and bias:** the crisis emphasized that accurate, reliable, and unbiased data models are imperative to ensuring accurate products, pricing, and in estimating the degree of risk. Again, AI mimics these concerns insofar as inaccurate data poses a threat to consumers, as well as the integrity of businesses individually and the industry as a whole.
- **Sufficient oversight:** naturally, insufficient oversight of the financial services industry, its products, and compliance attempts contributed to the crisis. In considering the adoption of AI, it is submitted that sufficient regulatory oversight and understanding is required to mitigate the manifestation of the aforementioned risks; this, however, relies on sufficient transparency and proper data and models being in place.
- **Coordinated approach:** prior to the crisis, legislation and regulatory efforts were insufficiently cohesive among financial services sectors and across nations; insofar as the industry operates across borders, this lack of coordination exposed systemic risks and complicated response efforts. Again, AI is inherently cross-border;

¹² FCA Handbook, SYSC 8.1, available at: SYSC 8.1 General outsourcing requirements – FCA Handbook, <https://tinyurl.com/25zv69p>

¹³ E.U. AI Act, Article 14

¹⁴ About the FCA, <https://tinyurl.com/5t42dnu9>

¹⁵ See Section 6.1.3; EU AI Act, Article 52.

to this end, ensuring some degree of consistency and coordination in regulatory efforts should act to mitigate such shortcomings.

- **Adaptive:** put simply, the financial crisis highlighted that financial regulation was insufficiently responsive to changes within the industry, with this leading to regulatory gaps and shortcomings. Insofar as AI and AI integration are evolving rapidly, it is submitted that any regulation must be able to adapt and respond to practical, industry developments in order to minimize regulatory pitfalls.

The U.K. government has affirmed that it wants to adopt a “pro-innovation approach” to AI regulation. In essence, they propose focusing regulatory efforts in a targeted, context-

Regulatory Genome Project

It is generally accepted that coordination in regulatory efforts should be considered in formulating financial services policy; however, given the volume, complexity, and divergence in existing financial services regulation, this is a time-consuming and difficult process. The Regulatory Genome Project (RGP),¹⁶ as developed by Cambridge Judge Business School, aims to address this issue through its application of machine learning and AI.

To expand, RGP uses AI technology and machine learning to analyze and compare regulatory principles relating to financial services. Data relating to global financial services regulation is inputted into the AI system; after processing this data, the system is able to derive international principles and regulatory standards. This information is shared through a “common information structure”, which allows regulators to quickly and “easily benchmark different regulatory frameworks,” allowing them to prepare for innovative developments. In essence, this open information model simplifies the sharing of regulatory requirements and considerations among jurisdictions, thereby permitting for greater coordination, supporting effective supervision, and creating greater regulatory efficiencies.

specific, and coherent fashion that permits for “safe” innovation.¹⁷ The below summarizes the various policy statements and regulatory proposals as provided by the FCA and the U.K. government in respect of AI, highlighting how they align with and adopt the above suggestions.

4.2 U.K. government approach to AI policy

4.2.1 REGULATORY STRATEGY AND ATTITUDE

As noted, the U.K. government has committed to a “pro-innovation approach” in the regulation of AI and AI systems. In 2021, various government departments released the “National AI strategy” that set out the “ten-year plan” for ensuring the U.K.’s position as “a global AI superpower”.¹⁸ Essentially, the strategy inferred that the widespread implementation of AI systems was inevitable and, to ensure market competitiveness, the government needed to support this transition through well-crafted regulation. Recognizing the need for adaptable and robust rules, the proposal was underpinned by three overarching and strategic themes: (i) the need to promote investment and to plan for the long term, (ii) the need to capture the benefits of AI across all sectors and regions, and (iii) the need to ensure proper understanding and governance of AI systems.

Irrespective of this, the government recognized that implementing regulation should not be done until it has a proper and full understanding of the risks that such regulation seeks to address.¹⁹ As such, in 2022, the Science, Innovation, and Technology Committee was tasked with launching an inquiry to explore AI’s impact on society, economy, and regulation. The ongoing inquiry has received over 100 written submissions and 24 oral testimonies that will serve to guide and support the implementation of robust and appropriate AI governance frameworks.

4.2.2 REGULATION

In July 2022, the U.K. government proposed new regulations for AI use,²⁰ broadly aligning with the National Strategy. To expand, the proposal reaffirms that the government is “firmly pro-innovation” but recognizes that this needs to be balanced against a “pro-safety” approach in order to ensure the adoption of the technology and foster public trust. Notably, the proposal does not promote AI-specific laws or regulations;

¹⁶ <https://tinyurl.com/nrmaxeuf>

¹⁷ Letter from DSIT Secretary of State and the Economic Secretary to the Treasury and City Minister to the Financial Conduct Authority, <https://tinyurl.com/3cxum2f4>

¹⁸ Guidance, “National AI strategy,” updated December 18, 2022, <https://tinyurl.com/ye22avk7>

¹⁹ As noted in Policy paper, “Establishing a pro-innovation approach to regulating AI,” July 20, 2022, <https://tinyurl.com/42hf8c86>

²⁰ *Ibid.*

instead, it focuses on core principles that are to apply across all industries. These principles address the key risks attributed to AI systems, focusing on safety, transparency, fairness, accountability, and contestability. Irrespective of this, the specific implementation of such principles is subject to the discretion of the industry regulator (so, for the purposes of financial services, the FCA). In this sense, the proposed regulation appears to strike a balance between adaptability and robustness: it addresses the key risks attributable to AI generally while retaining sufficient flexibility to address industry specific concerns.

4.3 FCA comment on AI policy

The FCA acts to regulate and supervise the conduct of financial services firms within the U.K. In doing so, it determines appropriate rules and guidance applicable to financial services businesses and the industry more generally; accordingly, the FCA will be the body responsible for the specific implementation of the proposed principles governing AI in respect of financial services, as discussed above.

Together with the BoE and Prudential Regulation Authority (PRA) (collectively the “supervisory authority”), the FCA published DP5/22;²¹ this report considered the specific application of AI regulation within the context of financial services, calling on industry participants to respond on issues including the degree and type of regulation. The report identified key risks relating to the integration of AI within financial services, including, but not limited to those of consumer protection risks and data concerns.

Once received, the industry responses and feedback were summarized in FS2/23.²² Notably, many respondents were not in favor of sector-specific definition for AI given concerns of rapid technological advancements and regulatory arbitrage. Some respondents suggested AI-specific rules were unnecessary altogether. Further, it was suggested that greater national and international coordination was required to mitigate industry fragmentation. Broadly, these considerations align with the proposed policy considerations set out above.

Although the regulators continue to formulate regulatory standards, it can be concluded that financial services institutions should prepare for incoming AI regulations and look to align themselves with the guiding principles.

5. INDIRECT REGULATION

In some instances, the application of AI in financial services will not generate any novel risks or regulatory concerns; here, such risks can be addressed through legislation and regulatory provisions that would otherwise apply to the financial services industry and institutions. The following will demonstrate how the application of AI in financial services can effectively fall within existing regulations.

5.1 Consumer protection

As noted, AI can be utilized to identify consumers by virtue of specified characteristics; in doing so, firms can tailor their products and services to better support the consumer and their specific needs. For example, this application permits for the identification of vulnerable persons who may need additional support or be more susceptible to malicious activity. However, through such identification, consumers are at a heightened risk of exploitation, bias, and discrimination. Such technology may be used in respect of adjustable-rate mortgages (ARM); the application of such AI systems in ARM-monitoring puts consumers at risk of predatory lending practices and unfair treatment, which could serve to exacerbate inequalities or financial vulnerabilities. Such risks may manifest due to insufficient datasets or the programming and personalization of the technology.

While many firms have voluntarily implemented policies and procedures to address such concerns,²³ they will likely be subject to the FCA’s Principles for Business (“the Principles”) and,²⁴ when implemented, its policy of “A New Consumer Duty” (“the Duty”).²⁵

- The Principles are fundamental obligations placed on firms to protect customers and, in particular, retail customers. For example, firms are under an obligation to pay due regard to customers interest and treat them fairly, and they must act to deliver good outcome for retail customers. More generally, the Principles serve to heighten protections (particularly for vulnerable customers) and mitigate the risk of discrimination. While not specific to AI, the Principles place a general duty on regulated firms operating within the financial services industry. Further, such Principles will also apply to AlaaS when the third-party service provider interacts with the

²¹ Bank of England, 2022, “Artificial intelligence and machine learning,” DP5/22, <https://tinyurl.com/47xds9dh>

²² Bank of England, 2023, “Response paper on artificial intelligence and machine learning,” FS2/23, October 26, <https://tinyurl.com/5bsua5b9>

²³ Bank of England (n 21)

²⁴ PRIN 2.1 The Principles – FCA Handbook, <https://tinyurl.com/4uh8yuh5>

²⁵ PS22/9: A new Consumer Duty | FCA, <https://tinyurl.com/bdev8k78>

regulated business. As such, and insofar as the Principles are sufficiently broad, they will mitigate the risks in this specific context.

- The Principles are supported further by the FCA's Vulnerable Customer Guidance.²⁶ In practice, these complement the Principles and inform firm's behavior in complying with their obligations in respect of vulnerable persons.
- The Duty serves to increase the responsibilities inferred on firms under the Principles; in essence, it requires that firms have a greater responsibility and "more positive role in delivering good outcomes for [retail] consumers" beyond their clients.²⁷

Further, legislation such as the Equality Act 2010 will apply to prohibit instances of discrimination; the Vulnerable Customer Guidance expressly notes that firms should have regard to the 2010 Act and aims to implement similar outcomes to the anticipatory duty on reasonable adjustments. Many of the protected characteristics overlap between the Guidance and 2010 Act, meaning that a breach of one will likely result in a concurrent breach of the other.

5.2 Data processing

In practice, AI systems will process significant quantities of data when fulfilling the set functions. Such data may, and likely will, include "personal data" as defined by Regulation (EU) 2016/679 (the 'GDPR'). Where personal data is processed as part of the activities of an E.U. entity, it must be done in accordance with the GDPR;²⁸ in essence, the data processor must have a lawful basis for the processing of such data and it must implement proper procedures whereby the data subjects can exercise their rights.

The primary question centers on whom assumes the position of (and liability as) the data processor. In theory, the AI system could be considered to be the data processor insofar as it is responsible for processing such data. However, and as discussed above, AI does not have a separate legal personality and so cannot assume the responsibilities attributable to a data processor under the GDPR. Thus, the issue centers on whether the data processor will be the AI provider, developer,

or the financial services organization adopting and utilizing the technology. In practice, this will be determined on a case-by-case basis. For example, where an organization opts for AlaaS the underlying service provider will likely be considered the data processor; on the other hand, where an organization has developed an in-house AI system, they will be considered the responsible party.

Irrespective of this, the principles and regulations within the GDPR will be applicable in this context. The factual circumstances and underlying risk remain the same; assuming the data processor can be properly identified, then the GDPR should prove efficient in addressing the issue of AI data processing.

6. DIRECT REGULATION

6.1 The E.U. AI Act

6.1.1 OVERVIEW AND APPLICATION

The E.U. is leading the way by being the first regulatory body to attempt to regulate AI, having approved a set of regulations to be applied to AI systems across Europe in early December 2023. The new rules are to be contained in the E.U. AI Act ("the Act"),²⁹ which is slated to take effect in early 2024. It will have a broad application, applying horizontally across all sectors; additionally, it has been attributed extra-territorial effect, so will apply to any third-country providers and users of AI systems where such systems or generated output is used within the bounds of the E.U. In essence, it aims to unify and coordinate regulatory efforts across member states while minimizing the risks attributed to AI systems within the context of the E.U.

6.1.2 A RISK-BASED APPROACH

The Act adopts a risk-based approach, focusing on addressing and regulating AI systems that present the greatest "risk" while simultaneously clarifying the obligations of the AI providers and deployers. To expand, it categorizes AI systems according to risk, with more stringent regulations being applied to those that present the most significant risks to E.U. persons and values. In this sense, the Act applies to AI systems generally instead of creating rules for specific industry sectors.

²⁶ FG21/1: Guidance for firms on the fair treatment of vulnerable customers, <https://tinyurl.com/23v5yw47>

²⁷ Bank of England (n 21) at 4.9

²⁸ See the Data Protection Act 2018 for the U.K. transposition.

²⁹ Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 8115/21, January 22, 2024.

It prohibits the categories of AI systems that are taken to present the greatest risk of causing harm; this includes exploitative and certain types of biometric identification system (e.g., emotion recognition and social scoring in various circumstances).³⁰ Such AI systems are deemed to create an “unacceptable” degree of risk insofar as they contravene E.U. values or constitute a sufficient threat to established fundamental rights. For example, developers and providers will not be able to put AI systems that would exploit specific vulnerabilities where the purpose of such exploitation is to materially distort the behavior of that person or group in a way that may cause significant harm on the market. Naturally, this acts to protect consumers insofar as financial services firms will not have access to such systems within the E.U. This may limit financial services firms from adopting such systems through AlaaS or external routes.

The Act also purports to limit the use of “high-risk AI systems” to narrowly defined instances that are subject to strict requirements.³¹ Such AI systems will be deemed “high-risk” where they present “significant potential harm” to E.U. persons and their “health, safety, fundamental rights” or, more broadly, the “environment, democracy and[/or] the rule of law.” In principle, it has been argued that the criteria adopted is sufficiently broad so as to encompass AI systems used to evaluate creditworthiness, grant loans, or facilitate other financial services activities. Accordingly, those who adopt such systems may need to adhere to the heightened obligations and regulatory burdens prescribed by the Act.

Irrespective of this, the E.U. has recognized that the test is sufficiently broad in its scope. As such, and to address borderline cases or potential compliance issues, providers must complete assessment documentation and registration documentation in an E.U. database before introducing the system to the E.U. market; the Commission will then determine whether the system presents a “high-risk” or would fall within a lower-risk category (as discussed below).

Where a system presents a “limited risk”, the provider must still adhere to some compliance requirements, although they are less onerous than those attached to high-risk systems. Essentially, such providers will be required to inform users that the content or system is AI generated. Where AI presents an even lower risk, providers are not obligated to adhere to any compliance efforts; rather, they are simply encouraged to implement voluntary codes of conduct and practice.

6.1.3 SYSTEMIC RISK

As noted, issues of systemic risk are addressed in the regulations addressing general-purpose AI (GPAI) models;³² this essentially refers to AI systems that show “significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is placed.”³³ A GPAI model will attach systemic risk where it has “high-impact capabilities”.³⁴ Providers of such models will be required to maintain up-to-date technical documentation and they must make any such information available to providers that integrate the AI in their systems.³⁵ Further, they must make information pertaining to the content used to train the AI system publicly available.³⁶ These obligations are accompanied by other monitoring and procedural requirements,³⁷ all of which address the concerns surrounding a lack of transparency and insufficient oversight. To this end, the Act addresses some of the primary risks attributable to the integration of AI systems in financial services, thereby removing barriers to its utilization.³⁸

6.1.4 RIGHTS AND OBLIGATIONS

Put simply, the Act distinguishes between the obligations borne by providers or developers and those borne by users. In practice, financial services firms are likely to be considered users rather than developers; however, it may be possible that a financial services firm becomes a developer should it develop its own AI system. Providers and developers must: ensure AI systems are transparent; inputted data is of a sufficient quality and integrity; they are accountable for the system; and that they comply with technical standards required by the E.U. Users must conduct proper risk assessments and comply with proper monitoring efforts.

³⁰ EU AI Act, Title II.

³¹ EU AI Act, Title III.

³² EU AI Act, Article 52.

³³ EU AI Act, Article 3(44b).

³⁴ EU AI Act, Article 52(1), as defined in Article 22.

³⁵ EU AI Act, Article 52c(1).

³⁶ EU AI Act, Article 52c(1)(d).

³⁷ EU AI Act, Articles 52d and 52e.

³⁸ See section 3.2.2.

The Act does not, in itself, create individual rights for those harmed by AI systems in practice;³⁹ rather, it does clarify and codify the obligations of the relevant parties. Further, it seeks to promote transparency within AI systems and their adoption within various industries. For financial services institutions, evidencing decision-making processes and justifying decisions will necessitate that they are transparent about their efforts and structures irrespective of whether they are the provider or user. As discussed, the inherent lack of certainty as to the allocation of liability and issues of transparency have presented the primary barriers for the adoption of AI in all industries; as such, it is submitted that the Act provides much needed clarity in support of AI integration.

6.2 The E.U. AI Liability Directive

The E.U. AI Liability Directive (“the Directive”)⁴⁰ aims to address potential claims for harm caused by AI systems. While at an earlier stage of the legislative process, it is intended to accompany the Act and the clearer obligations set out therein.

The Directive will apply to AI systems that are available to, or operate within, E.U. markets; in doing so, it shall act as a standard of minimum harmonization (i.e., persons may elect to invoke national laws where they appear more favorable), but will need to be transposed into national law. It seeks to address the shortcomings of traditional liability rules in addressing claims for harm against AI systems; as such, the proposals focus on addressing the difficulties of proof attaching to the complexities introduced by AI.⁴¹ In doing so, the Directive aims to recognize the nuances of AI and, therefore, sets out a new evidentiary mechanism; this mechanism aims to address the lack of transparency and complexity associated with AI systems. In doing so, the Directive also aims to establish a presumption of causation between the defendant and harm complained of.

Thus, when read with the Act, the proposed procedural rules could alleviate some of the key barrier to corporate integration and adoption of AI insofar as it purports to clarify the extent and allocation of liability; however, at the time of writing, it remains subject to EU approval and, therefore, has no binding legal effect.

7. ETHICAL AI

“Ethical AI” requires that AI systems are developed, implemented, and used in ways that align with ethical standards, respecting established values and fundamental human rights. In doing so, ethical AI seeks to advance the transformative potential of AI systems while protecting human values and societal wellbeing. Achieving this in practice requires robust guidelines, with industry participants and policymakers agreeing to set principles. It is a critical component of any organizational strategy on AI.

Validate AI, an independent community interest company, focuses on improving the validation of AI and have developed a number of whitepapers and voluntary codes of conduct to this end. The most recent whitepaper has been the subject of wide engagement, setting out a framework that supports the widespread adoption of ethical AI.⁴² To expand, the approach focuses on six fundamental pillars, with each addressing risks commonly associated with AI integration. The following sets out each of the pillars, highlighting how they serve the underlying aim of ethical AI:

- (i) **Responsibility and accountability:** organizations should be held accountable for the consequences of the systems they develop, with this being central to the degree of risk attaching to the product. Validate AI suggest that developers should appoint an AI officer responsible for monitoring risks and managing the responsible deployment of AI systems.
- (ii) **Code of practice:** codes of practice are central to ensuring that AI systems are deployed to certain standards; Validate AI submit that practitioner focused codes of conduct are required “to ensure that AI systems can be trusted.”
- (iii) **Convening:** convening and coordination are key to ensure all stakeholders are heard when considering the deployment and regulation of AI systems.
- (iv) **Independent audit:** audits are viewed as particularly useful where high-impact AI systems are at issue insofar as they act to mitigate the likelihood that inappropriate, high-risk systems are deployed. This is common practice in other industries where public safety concerns are relevant.

³⁹ Cf. Section 6.2.

⁴⁰ EUR-Lex (n 8)

⁴¹ See Section 3.2.1.

⁴² 14072023 Validate AI – Our position to tackling AI risk, <https://tinyurl.com/ya9zyzyuk>

(v) **Monitoring:** AI systems should be continuously monitored after deployment, with contingency plans in place to manage a number of scenarios. This educates relevant parties as to the nature of the specific AI system while providing protection against the risks of failure.

(vi) **Education:** educating industry participants, businesses, the general public, and governments about AI and the associated risks is key to ensuring those parties are able to properly assess and make informed decisions about AI systems they may interact with. Validate AI suggest that “education should be practitioner-centric,” ensuring that industry participants can apply ethical standards in their development roles. Similarly, they suggest that general education can be tailored to the application of AI in different industries.

Together, these pillars act to promote fundamental values and improve the social responsibility in the adoption of AI, thereby mitigating the aforementioned risks and removing barriers to the development and implementation.

8. CONCLUSION

AI systems are valuable tools that can be applied in nearly any industry; they are of particular utility in the context of financial services, where the management and use of data has been the foundation of businesses since their inception.

It is, however, clear that some degree of regulatory intervention is required to enable the most efficient integration of the technology. The proper application of public policy and the specifics of regulation remain uncertain. While obvious, the need to balance innovation with safety is difficult to strike. Alongside this, international competitiveness has become a critical focus for policymakers and remains a significant challenge for businesses (particularly those that are cross-border in nature). However, financial services firms are familiar with these high-level questions and challenges; businesses are demonstrating an increased understanding of the benefits to be derived from AI systems and through engaging with fintech partners, suggesting these barriers are not insurmountable; from this, it is apparent that the adoption of industrial data processing and the use of novel AI systems will continue among the most successful financial services firms over the coming years.

© 2024 The Capital Markets Company (UK) Limited. All rights reserved.

This document was produced for information purposes only and is for the exclusive use of the recipient.

This publication has been prepared for general guidance purposes, and is indicative and subject to change. It does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (whether express or implied) is given as to the accuracy or completeness of the information contained in this publication and The Capital Markets Company BVBA and its affiliated companies globally (collectively "Capco") does not, to the extent permissible by law, assume any liability or duty of care for any consequences of the acts or omissions of those relying on information contained in this publication, or for any decision taken based upon it.

ABOUT CAPCO

Capco, a Wipro company, is a global technology and management consultancy focused in the financial services industry. Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on Facebook, YouTube, LinkedIn and Instagram.

WORLDWIDE OFFICES

APAC

Bengaluru – Electronic City
Bengaluru – Sarjapur Road
Bangkok
Chennai
Gurugram
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

MIDDLE EAST

Dubai

EUROPE

Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
Glasgow
London
Milan
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA

Charlotte
Chicago
Dallas
Hartford
Houston
New York
Orlando
Toronto

SOUTH AMERICA

São Paulo

WWW.CAPCO.COM



CAPCO

a wipro company