

WORLD
OF
DATA
GENAI

a wipro company

THE CAPCO INSTITUTE
JOURNAL
OF FINANCIAL TRANSFORMATION

TECHNOLOGY

Multimodal artificial intelligence:
Creating strategic value
from data diversity

CRISTIÁN BRAVO



GenAI

2024/2025 EDITION

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 Birchler, 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

TECHNOLOGY

08 Mindful use of AI: A practical approach

Magnus Westerlund, Principal Lecturer in Information Technology and Director of the Laboratory for Trustworthy AI, Arcada University of Applied Sciences, Helsinki, Finland

Elisabeth Hildt, Affiliated Professor, Arcada University of Applied Sciences, Helsinki, Finland, and Professor of Philosophy and Director of the Center for the Study of Ethics in the Professions, Illinois Institute of Technology, Chicago, USA

Apostolos C. Tsolakis, Senior Project Manager, Q-PLAN International Advisors PC, Thessaloniki, Greece

Roberto V. Zicari, Affiliated Professor, Arcada University of Applied Sciences, Helsinki, Finland

14 Understanding the implications of advanced AI on financial markets

Michael P. Wellman, Lynn A. Conway Collegiate Professor of Computer Science and Engineering University of Michigan, Ann Arbor

20 Auditing GenAI systems: Ensuring responsible deployment

David S. Krause, Emeritus Associate Professor of Finance, Marquette University

Eric P. Krause, PhD Candidate – Accounting, Bentley University

28 Innovating with intelligence: Open-source Large Language Models for secure system transformation

Gerhardt Scriven, Executive Director, Capco

Tony Moenicke, Senior Consultant, Capco

Sebastian Ehrig, Senior Consultant, Capco

38 Multimodal artificial intelligence: Creating strategic value from data diversity

Cristián Bravo, Professor, Canada Research Chair in Banking and Insurance Analytics, Department of Statistical and Actuarial Sciences, Western University

46 GenAI and robotics: Reshaping the future of work and leadership

Natalie A. Pierce, Partner and Chair of the Employment and Labor Group, Gunderson Dettmer

ORGANIZATION

56 How corporate boards must approach AI governance

Arun Sundararajan, Harold Price Professor of Entrepreneurship and Director of the Fubon Center for Technology, Business, and Innovation, Stern School of Business, New York University

66 Transforming organizations through AI: Emerging strategies for navigating the future of business

Feng Li, Associate Dean for Research and Innovation and Chair of Information Management, Bayes Business School (formerly Cass), City St George's, University of London

Harvey Lewis, Partner, Ernst & Young (EY), London

74 The challenges of AI and GenAI use in the public sector

Albert Sanchez-Graells, Professor of Economic Law, University of Bristol Law School

78 AI safety and the value preservation imperative

Sean Lyons, Author of Corporate Defense and the Value Preservation Imperative: Bulletproof Your Corporate Defense Program

92 Generative AI technology blueprint: Architecting the future of AI-infused solutions

Charlotte Byrne, Managing Principal, Capco

Thomas Hill, Principal Consultant, Capco

96 Unlocking AI's potential through metacognition in decision making

Sean McMinn, Director of Center for Educational Innovation, Hong Kong University of Science and Technology

Joon Nak Choi, Advisor to the MSc in Business Analytics and Adjunct Associate Professor, Hong Kong University of Science and Technology

REGULATION

104 Mapping GenAI regulation in finance and bridging the gaps

Nydia Remolina, Assistant Professor of Law, and Fintech Track Lead, SMU Centre for AI and Data Governance, Singapore Management University

112 Board decision making in the age of AI: Ownership and trust

Katja Langenbucher, Professor of Civil Law, Commercial Law, and Banking Law, Goethe University Frankfurt

122 The transformative power of AI in the legal sector: Balancing innovation, strategy, and human skills

Eugenia Navarro, Lecturer and Director of the Legal Operations and Legal Tech Course, ESADE

129 Remuneration on the management board in financial institutions: Current developments in the framework of supervisory law, labor law, behavioral economics and practice

Julia Redenius-Hövermann, Professor of Civil Law and Corporate Law and Director of the Corporate Governance Institute (CGI) and the Frankfurt Competence Centre for German and Global Regulation (FCCR), Frankfurt School of Finance and Management

Lars Hinrichs, Partner at Deloitte Legal Rechtsanwaltsgesellschaft mbH (Deloitte Legal) and Lecturer, Frankfurt School of Finance and Management



CAPCO CEO WELCOME

DEAR READER,

Welcome to our very special 60th edition of the Capco Journal of Financial Transformation.

The release of this milestone edition, focused on GenAI, reinforces Capco's enduring role in leading conversations at the cutting edge of innovation, and driving the trends shaping the financial services sector.

There is no doubt that GenAI is revolutionizing industries and rapidly accelerating innovation, with the potential to fundamentally reshape how we identify and capitalize on opportunities for transformation.

At Capco, we are embracing an AI infused future today, leveraging the power of GenAI to increase efficiency, innovation and speed to market while ensuring that this technology is used in a pragmatic, secure, and responsible way.

In this edition of the Capco Journal, we are excited to share the expert insights of distinguished contributors across academia and the financial services industry, in addition to drawing on the practical experiences from Capco's industry, consulting, and technology SMEs.

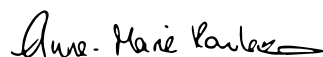
The authors in this edition offer fresh perspectives on the mindful use of GenAI and the implications of advanced GenAI on financial markets, in addition to providing practical and safe frameworks for boards and firms on how to approach GenAI governance.

The latest advancements in this rapidly evolving space demonstrate that the potential of GenAI goes beyond automating and augmenting tasks, to truly helping organizations redefine their business models, processes and workforce strategies. To unlock these benefits of GenAI, I believe that firms need a culture that encourages responsible experimentation and continuous learning across their organization, while assessing the impact of the potential benefits against a strategic approach and GenAI framework.

I am proud that Capco today remains committed to our culture of entrepreneurialism and innovation, harnessed in the foundation of our domain expertise across our global teams. I am proud that we remain committed to our mission to actively push boundaries, championing the ideas that are shaping the future of our industry, and making a genuine difference for our clients and customers – all while ensuring to lead with a strategy that puts sustained growth, integrity and security at the forefront of what we do.

I hope you'll find the articles in this edition both thought-provoking and valuable as you create your organization's GenAI strategy and future direction. As we navigate this journey together, now is the time to be bold, think big, and explore the possibilities.

My greatest thanks and appreciation to our contributors, readers, clients, and teams.



Annie Rowland, **Capco CEO**

MULTIMODAL ARTIFICIAL INTELLIGENCE: CREATING STRATEGIC VALUE FROM DATA DIVERSITY

CRISTIÁN BRAVO | Professor, Canada Research Chair in Banking and Insurance Analytics, Department of Statistical and Actuarial Sciences, Western University¹

ABSTRACT

The modern revolution of artificial intelligence (AI) has a benefit that is often not mentioned: it allows the use of diverse data from multiple sources and of multiple types (multimodal data), such as video, audio, or images, in an efficient, and, more importantly, effective manner. While this is much closer to how experts make decisions, the challenges are that it must be done profitably, while considering the internal culture and the operational systems that are available to ensure a positive return on investment (RoI). In this article, I will summarize some of the advantages and point out some of the challenges in creating effective and useful AI systems that leverage multimodal data.

1. INTRODUCTION

If you have been doing something for a long time, you probably use some sort of multimodality to make your decisions. To start this discussion, let us imagine a financial institution deciding on whether to underwrite a large bond placement in the market. An internal group of analysts will study the financial situation of the company (structured data), read reports regarding the market (text data), maybe listen to the last investor call (audio data), watch the last interview by the CEO on the local business news channel (video data), talk to experts, and analyze a long list of data sources that will help them decide if underwriting the operation is a good idea. Among these data sources, there will most likely be some scores and ratings that come from models. Maybe the analysts will use ChatGPT or other large language model (LLM) to summarize reports, but the final decision will come from interpreting all these data sources and joining them in some sort of mental or world model to arrive at a conclusion.

Hence, complex decision making is multimodal; why are our models not? This is not a capricious statement. When an expert makes a decision, it is made by combining past experiences with many sources of information in a complex, integrated manner. In these cases, AI can be a support or a hinderance. A few studies have categorically shown this. De-Arteaga et al. (2020) show that expert workers are more likely to override automated systems when the recommendations they provide go against their knowledge and experience. Lebovitz et al. (2022) find that when the AI systems that provide medical recommendations are opaque, experts will resist accepting them and fail to integrate the models into their processes. On the other hand, van den Broek et al. (2021) find that when the systems are perceived as useful, experts reach a hybrid process that enriches their knowledge with the recommendations by the AI system.

Most likely, if you work at a financial institution or a fintech company, you already have some sort of multimodal model deployed or interact with one from a provider regularly. It is common for banking apps to offer check deposits with photos,

¹ I acknowledge the support of the Canada Research Chairs Program [CRC-2018-00082].

for example. This requires the model to understand handwritten characters, connect to the structured data of the app and the customer's account information, match the information on the check with the account number, and most likely consider the response of a fraud detection model that receives the data and decides whether the deposit should be accepted or not. This can be perceived as a simple operation, but the modeling behind it is anything but. In the next couple of sections, I will highlight the advantages and challenges of deploying a multimodal model. These come from my own experience in developing these models and supporting institutions in deploying them, using, for example, text and numerical data [Stevenson et al. (2021)], LiDAR data together with sociodemographic information [Stevenson et al. (2022)], social network data in combination with behavioral data for credit risk [Zandi et al. (2024)], or combining financial information and time series market data for portfolio optimization [Korangi et al. (2024)], to name a few. Together with two colleagues, I am currently in the process of writing a book that covers the technical details of developing multimodal models in the financial services sector [Deep learning in banking, Wiley, forthcoming 2025], and I invite you to have a look if you are curious about this topic beyond this short article.

2. AI MULTIMODAL MODELS

Starting with the basics: a multimodal model is any model that uses more than one “modality” (type) of data to generate an output. A multimodal model can be a simple model that, for example, takes raw text (one, unstructured, modality) and structured information about who wrote the text (another, structured, modality), counts the number of sad emojis versus happy emojis and decides if the text has a positive or negative sentiment. Of course, this is probably a terrible model. Most modern models use some sort of deep learning AI architecture, in particular a “transformer” architecture [Vaswani et al. (2017)], to generate outputs. The transformer architecture is, in overly simple words, a statistical model that takes sequence-like data (such as text, audio, time series, and many others) and does a series of numerical processes (multiplying matrices) to generate features that describe a given outcome. This starts with generating an “embedding”, or a numerical representation of the sequence data. After what can be a massive series of calculations, the original representation is transformed into an output. This can be, for example, a probability, a forecast, or an embedding of the next word in the sequence. This latter example is the basis

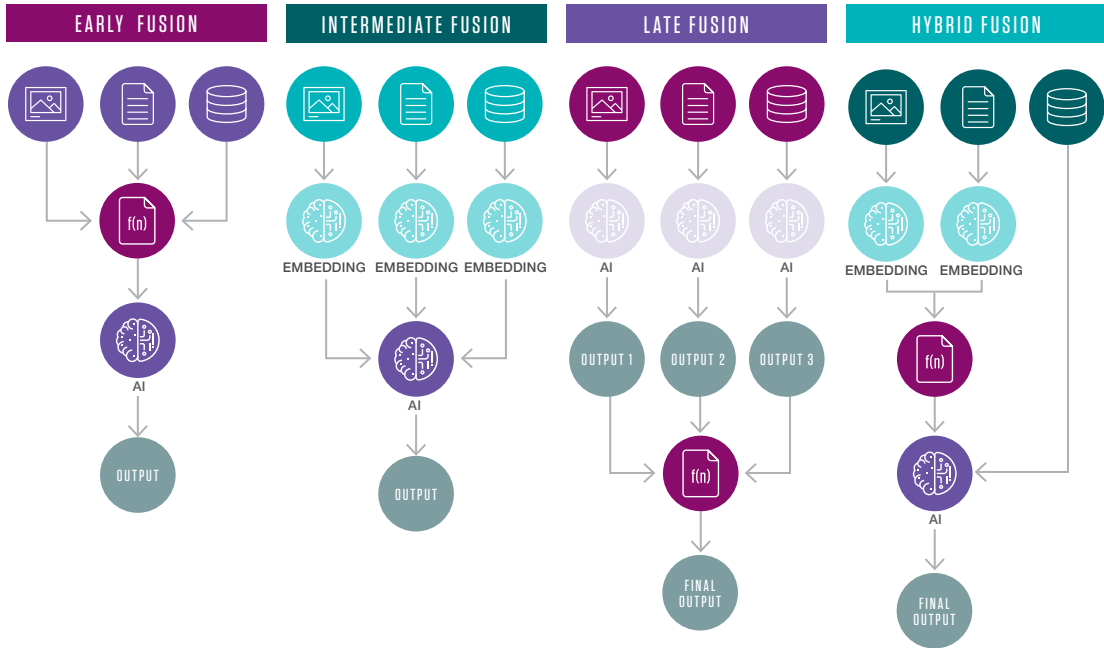
of the modern LLM systems. They take a text sequence and, through a series of transformers, predict the next word in the sequence. They do this iteratively until it predicts the sentence has reached its end.

To make a model truly multimodal, we must include many data sources. These are then combined using a data fusion strategy. We can do this at the initial step (early fusion), after it has been processed by an initial model (intermediate fusion), or even after it has been fully processed, by combining different models' outputs (late fusion). Figure 1 illustrates these fusion methods. It is also very common, and even desirable, to use more than one fusion strategy. A popular one includes using early fusion to combine similar data types, for example all numerical data modalities, and later using intermediate fusion to combine these numerical and unstructured data embeddings before processing them for prediction by dense layers. We call this process “hybrid fusion”, and it is currently the most common methodology applied in multimodal models.

The structure in Figure 1 does not showcase what exactly the embedding model, the AI model, and the output model are. This will vary depending on the application and the complexity of the model being deployed. Some examples are using featurization and then concatenation. Featurization simply means extracting numerical variables from the data, such as counting the emojis in our example above, but can become much more sophisticated. By concatenation, we mean that once the features are calculated, we create a unique numerical representation by joining the outputs of the individual features together. More sophisticated strategies use models to create these combinations. For example, in Tavakoli et al. (2023), we use Cross Attention, a type of deep learning architecture that has been shown to be useful when fusing information (J. Zhang et al. 2023). This method works by combining two or more data representations using a specialized transformer. Furthermore, lately some research has been done on using LLM themselves as the fusion strategy (Zhang et al. 2024), where now the LLM processes the output of an embedding step and tries to combine it into a fused embedding by interpreting as language.

For the embedding step, there are also many choices. The most common methodology is to use a model, tailored to that modality, that can generate a numerical output. In text, for example, transformer-based models can be used to obtain them. The package “sentence-transformer” [Ubiquitous Knowledge Processing Lab (2024)] is the best known one,

Figure 1: Different information fusion strategies



In this diagram, adapted from Tavakoli et al. (2023), there are three modalities (structured data, text, and images) and they are either processed by an aggregation function, as in the early fusion example, or they are processed by a model that outputs an embedding, a numerical representation of the data. A final AI model then transforms these modalities into a desired final output, such as a prediction or a regression value.

having a considerable library of contributed models by the likes of Google, Nvidia, and many others. For image and video data, options are less plentiful. Google has a few options in their cloud services, as part of their MediaPipe offering [Google (2024)], as does Amazon within AWS with their Titan models [Amazon AWS (2023)]. Both companies also offer direct multimodal embeddings, which use an information fusion strategy to provide a unique embedding vector that can then be fed to the output layer.

After a multimodal embedding has been created, the last step is to generate an output. This depends on the task. An AI generative model will use a series of decoders that will predict what the next embedding in a sequence is. For a prediction or regression task, it is common to create a dense neural network that generates an output in a desired format, either a probability or a regressor. These can get more exotic depending on the application. The key takeaway is that a multimodal model is a flow of smaller tasks that work together via information fusion to create a larger, complex, representation of the data. This is then fed to a final model that will construct whatever output we require.

3. CREATING A MULTIMODAL MODEL

While there are several frameworks to train a model, the objective of this article is to discuss the strategic issues and challenges that arise when planning one strategically. For the technical discussion on training these models, I refer the reader to the many online guides by vendors, to the many online resources available, or to my own previously cited works. However, deciding first if one of these models is needed is the core issue I want to start with.

3.1 Defining the problem and identifying the data

Let us start by thinking about whether there is even a need to deploy a model using multimodal data. There are three questions that a savvy manager can ask themselves:

1. Is there a problem that we have not been able to solve with traditional models?
2. Do we have data in our data lakes/data warehouses that is not being used or is underutilized?
3. What is our current technical capacity?

Figure 2: A simple project prioritization matrix

	Low Rol	High Rol
High risk	No-go	Opportunities
Low risk	Back burner	Quick wins

Quick wins are models that should take priority. Opportunities are difficult models that can create “moats”. Back burners are projects that can be developed if there is spare capacity. No-go are projects with low ROI that are also very risky to develop.

Within different organizations, there will most likely be a mix of answers to these questions. Let us imagine a traditional bank that has been collecting the social media mentions of their SME customers. This information is used in their marketing propensity models by simply counting the number of mentions in a 90-day period, to identify if the SME is generating buzz and may have a need for funds the bank can provide. This information is combined with financial transaction and a simple regression model generates a propensity probability. Starting with the first question, the answer may be that the model fails to identify negative buzz, and thus it is generating incorrect offers to companies that are in a downturn or subject to media controversy. The second question follows, and the answer is that we are indeed underutilizing the text data in the data lake as we are only generating this buzz indicator, without considering the context that a more sophisticated model can bring. The third question is much more complex to answer. If the managers of our example institution desire to move forward with a more complex multimodal model, they will need to identify if they have the correct collaborators who can develop the model in their data science teams, and whether there is on-premises or cloud infrastructure that can be used to develop the models. Cloud GPU infrastructure can get expensive fast, and on-premises infrastructure may not be sufficient to train models, only to deploy them.

With the answer to these questions, an Rol analysis must be performed to identify the data, and the training and deployment costs are balanced to merit moving forward with the model. It has been famously said that 50% of data science models fail. This is, in large part, due to not identifying which models have good Rol and which ones do not. A strategy that has worked well for my own collaborations with corporate partners has been to characterize them on a simple matrix depending on risk of development failure versus Rol, as in Figure 2.

In this prioritization, which admittedly requires significant knowledge of the organization to correctly utilize, it is easy to identify which models have the highest priority. Low risk/High Rol models (Quick wins) are the ones that should take

precedence as they are most likely easily developed, deployed, and will have a smaller risk of failure. However, these are also models that competing companies with a similar sophistication level can develop just as easily. They will most likely become commonplace very soon.

The second quadrant is far more interesting, excusing the bias that we academics have for shiny new solutions. These are far more complex models to develop, which means they cannot be easily copied by direct competitors. A successful project from this quadrant corresponds to what now is famously referred to as a “moat”. A development that gives a competitive advantage. If a company can successfully develop a model in the “opportunity” quadrant, they will have something that is challenging to replicate and provides high return. However, the risk of failure here is much higher, so strong leadership is a must.

The other two quadrants are important because projects that fit in them must be identified, and resources will most likely be better used elsewhere. It is difficult to acknowledge that an idea that one cherishes is a back burner, meaning its difficulty is low, but its return is also comparatively low against other ideas. These are the projects that can be developed whenever there is spare capacity, but if given priority, they will result in low impact and can cause more harm than good. The final quadrant, “no-go”, includes models with low potential return and high risk. These models should not be developed, better alternatives certainly exist.

How to determine Rol? This has been acknowledged as a difficult challenge [PwC (2024)]. Some factors that involve Rol can be seen in Figure 3. For the risk of development, this will mean balancing what we identified before: technical skills of the company, current computational resources available, and the ever-important cultural challenge of changing or intervening a process. The last one is one of the most significant ones in modern generative AI (GenAI). It can cause rejection within teams or customers if it is quasi-human, but clearly robotic (the uncanny valley effect), or if there is fear that it will replace jobs and should be sabotaged. This has to be managed and monitored internally, providing proper training on the use of the tool, and reassurance of its purpose within the organization.

If a model is flagged for development, then the data collection must occur. This should be done by the organization’s data engineers. Cleaning it and leaving it in a state that can be used and generating ETL pipelines that can be deployed must be accounted for when calculating the project’s Rol and risk. Next comes model training.

3.2 Training the model

At this stage, the decision made in the previous step on the strategy to train the model must be followed. The financial services sector, and most non-tech companies for that matter, have the challenge that they do not normally have the computational resources necessary to train the model, but most of the time they do have the capacity to deploy the model. Here is where scalable cloud infrastructure helps. Normally, training is far more resource-intensive than deploying a model (except for some specific high-frequency models), so a cloud solution may be best, unless the organization has significant computational resources available. Most of the time this happens, it is because the organization has mature data science teams that are constantly piloting new models. In this case, on-premises training infrastructure makes sense, as cloud costs can quickly skyrocket. This is especially true for multimodal models, where their very nature makes them far more resource-intensive than their unimodal counterparts.

Training a multimodal model is a challenge for even the most sophisticated institutions. For example, the model we developed in Korangi et al. (2024) took well over two weeks to train on a distributed system with tens of modern GPU units, and I would consider this a relatively modest multimodal problem. Cloud training of a similar model using spot Amazon

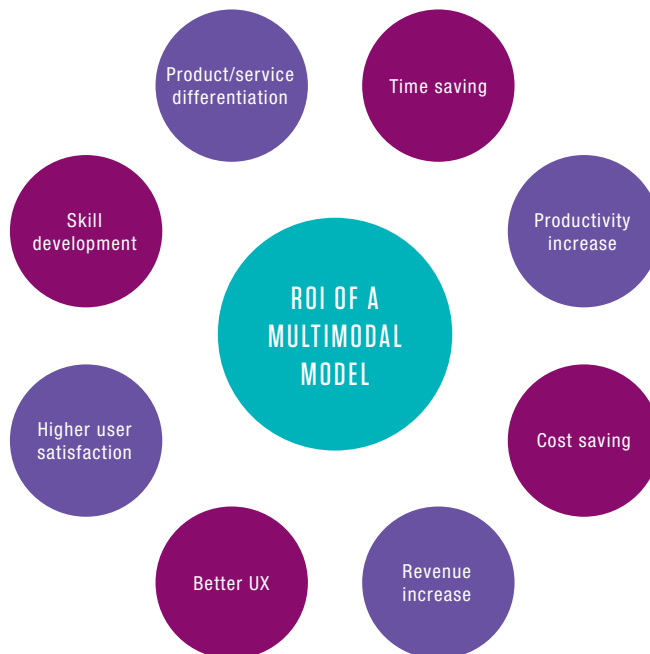
AWS instances would have run well into the six figures. Again, these calculations should have arisen from the RoI calculations in the previous step.

To effectively train the models, careful management must be followed. I can refer the interested reader to our work in this area, such as the previously referred Korangi et al. (2024) in portfolio optimization, Tiukhova et al. (2024) on credit card referral marketing, Zandi et al. (2024) and Tavakoli et al. (2023) on credit risk management, or the always rising literature discussing these deployments that can be found online. Suffice it to say, multimodality requires sufficient data science expertise internally. Subject matter experts are a requisite to train these models effectively, and whether the organization has this expertise should have been analyzed before the project began.

3.3 Deploying multimodal models

Once the model is created, the next step is to deploy it effectively. This can be done either on-premises, on cloud, or an edge deployment on the user's devices. The choice to do so must come from carefully evaluating the capacity of the company and their customers' needs, but some guidelines exist. It will depend on the type of model, the frequency it will be required, the types of devices the model will be served on, and the data security required, among other factors.

Figure 3: Some RoI factors



To begin with, the type of delivery will have an impact. I assume that some sort of large multimodal model (LMM) is being delivered, with a billion+ parameters. Most models can be “ablated”, a process to eliminate some redundant weights, to be reduced in size after training. They can also be “quantized” to reduce their size even further, by representing them in lower numerical precision. These techniques must be applied and the smallest, sufficiently performant, model must be the candidate to be deployed.

To consider if deployment must happen on premises, on the cloud, or on the device, we can start by deciding who will access the model and with which frequency. Worldwide customers frequently using a complex model at random intervals, suggests on-cloud deployment is best, as the scalable, distributed, nature of the cloud can help serve customers better. If the model is used by customers accessing their private data, our cloud choices may be limited or even forbidden by regulation, making an on-premises or edge deployment mandatory. Regarding this latter example, edge deployment applies to smaller models, but their usefulness is growing. Very recently, Microsoft developed a very aggressive 1-bit quantization that can deploy models directly on the customer’s devices [Ma et al. (2024)]. This technology is extremely recent though, but I expect it will become much more mainstream. A large range of options exist here, and the decision between edge, cloud, and local deployments must be carefully considered.

A second factor is infrastructure availability. Does your organization have sufficient capacity to deploy models internally? Servers, bandwidth, IT experts, cooling, and all the different parts of modern data centers must be available for on-premises deployment. This is most likely available in larger organizations, but not so much with startups and smaller fintech companies. Ongoing costs and growth strategies must be balanced to decide if on-premises is a sustainable strategy.

And finally, a model validation and monitoring plan must be in place to measure the performance of the model. At the pilot stage, I am a firm believer that A/B tests with carefully designed measures according to the original business plans are paramount. Research has shown that companies that use A/B tests in their operations have a higher RoI and better organizational learning strategies than the ones that do not [Koning et al. (2022)]. After deployment, standard “machine learning ops” (MLOps) procedures must be followed to ensure the model performs as expected. One specific challenge to multimodal models is that their building blocks are subject

to constant improvements as new sub-models for those modalities appear, making the “continuous improvement” leg of MLOps even more important. For example, in October 2024, the company Mistral AI released a 3-billion parameter model called Ministral [Mistral AI (2024)] that they claim is better than their 2023 7-billion-parameter model. If a local multimodal model has this older model deployed, it may make financial and statistical sense to replace it by their smaller and higher performant variant. Continuous “integration, training, delivery and monitoring”, the four core components of MLOps, are even more significant in multimodality.

4. SOME SPECIFIC CHALLENGES IN THE FINANCIAL SERVICES SECTOR

To finish this discussion, I want to touch on the specific challenges that financial institutions face when deploying multimodal models. The first challenge is a common one that may be more prevalent for multimodal data: designing data pipelines around legacy systems. Many financial institutions, particularly banks, are still running legacy systems that sometimes are sources of multimodal data, such as call center recordings or SWIFT transactional records. Designing new solutions may come with unwanted or unplanned overheads, such as creating a COBOL codebase that can create a dataset needed for a multimodal model deployed using PyTorch Lightning over a cloud server. Such technological chimeras must be identified and their benefits and costs carefully balanced.

A second point, not exclusive but certainly key in financial services, is change management and the culture of the company and its customers. If this is an internal model, how it is presented and deployed internally is a significant issue. In a recent project, in the context of an internal model to support customer service agents, we realized there was significant resistance to the use of GenAI by the organization’s collaborators as it was suspected it would lead to staff reduction. This threatened the success of the project altogether. The solution was to introduce the project by showcasing how it would help them, and how the model was simply an aid, not a replacement. This human-in-the-loop approach was practical: transformer-based generative models hallucinate, so human supervision is paramount to ensure these mistakes are caught early. Once the users realized the model was there to help them and that they remained the core owners of the workflow, their opinion shifted immediately. The model became “empowering”.

The final key issue is the regulatory hurdles that are inherent to model deployment in financial services. Tackling regulation and model management is a problem about which many articles have been written. Depending on the organizational area the model is deployed to (risk, marketing, operations, or any other), there will most likely be a series of regulatory hurdles that must be tackled to deploy the model. No matter the area the model is meant to support, model governance will be vital. One of the core aspects of modern machine learning deployment is accountability, who is responsible for the model performance, usage, and monitoring. Properly defining the responsibilities of model management, and how these models fit within the regulatory requirements that the organization is subject to, will greatly improve the chances a model is successful.

5. CONCLUSION

This article discusses multimodality and how AI can now leverage multiple sources of data. If you are reading this and work at any modern financial institution, I am sure there is some multimodal data somewhere in the organization that now just generates storage costs but can become key in multimodal development.

The core issue in generating a multimodal model is to balance RoI with the risk of development and deployment. Multimodality is tricky, requiring very complex individual parts working together in tandem. For example, a text-image-structured model can have a small 3B parameter LLM to generate sentence embeddings, a vision transformer to generate image

embeddings, a dense neural network to process the structured data, a cross-attention transformer to generate multimodal fusion, and a series of dense layers to generate a prediction. This can make the model scale to the tens of billions of parameters with relative ease, so identifying the right RoI opportunities is key. However, a correctly designed multimodal model also creates a moat: it is challenging to replicate and difficult to develop by competitors, while also generating internal skills within the data science teams that will not be common in the marketplace. All these considerations must be balanced when green-lighting a multimodal AI development.

There are both generic and specific challenges that must be considered when developing and deploying models in financial institutions. The biggest generic challenge is culture change, as AI models, particularly generative ones, can cause significant resistance. The more specific challenges within the financial services sector are transparency and accountability requirements, regulatory oversight and the risk of being a first mover in regulated models, and whether some of the multimodal data comes from legacy systems. Identifying such risks, managing them, and mitigating them can be the key to avoiding failure in an otherwise technically sound deployment.

Multimodality is the near future of AI. LLMs are already evolving into “large multimodal models”. Questioning now whether current AI and machine learning developments can be enriched by multimodality, or if new multimodal models can be created that solve previously unsolvable problems, can bring competitive advantages arising from better leveraging the diversity of data modern financial institutions and their customers create.

REFERENCES

- Amazon AWS, 2023, "Amazon Titan Image Generator, multimodal embeddings, and text models are now available in Amazon bedrock," AWS News Blog, <https://tinyurl.com/yck5pbvh>
- De-Arteaga, M., R. Fogliato, and A. Chouldechova, 2020, "A case for humans-in-the-loop: decisions in the presence of erroneous algorithmic scores," Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, 1-12, <https://tinyurl.com/dx6t8dah>
- Google, 2024, "Image embedding task guide | Google AI edge," <https://tinyurl.com/4f53eyvw>
- Koning, R., S. Hasan, and A. Chatterji, 2022, "Experimentation and start-up performance: evidence from A/B testing," *Management Science* 68:9, 6434-6453
- Korangi, K., C. Mues, and C. Bravo, 2024, "Large-scale time-varying portfolio optimization using graph attention networks," <https://tinyurl.com/5mhxejhe>
- Lebovitz, S., H. Lifshitz-Assaf, and N. Levina, 2022, "To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis," *Organization Science* 33:1, 126-148
- Ma, S., et al., 2024, "The era of 1-Bit LLMs: all large language models are in 1.58 bits," <https://tinyurl.com/3d3fh7bv>
- Mistral AI, 2024, "Un Ministral, Des Ministraux," <https://tinyurl.com/2unef8ep>
- PwC, 2024, "Solving AI's ROI problem. It's not that easy," PricewaterhouseCoopers, <https://tinyurl.com/dmndjrw7>
- Stevenson, M., C. Mues, and C. Bravo, 2021, "The value of text for small business default prediction: a deep learning approach," *European Journal of Operational Research* 295:2, 758-771
- Stevenson, M., C. Mues, and C. Bravo, 2022, "Deep residential representations: using unsupervised learning to unlock elevation data for geo-demographic prediction," *ISPRS Journal of Photogrammetry and Remote Sensing* 187, 378-392
- Tavakoli, M., R. Chandra, F. Tian, and C. Bravo, 2023, "Multi-modal deep learning for credit rating prediction using text and numerical data streams," <https://tinyurl.com/393b84rt>
- Tiukhova, E., E. Penalzoza, M. Oskarsdottir, B. Baesens, M. Snoeck, and C. Bravo, 2024, "INFLECT-DGNN: influencer prediction with dynamic graph neural networks," *IEEE Access* 12, 115026 - 115041, <https://tinyurl.com/msdzu86>
- Ubiquitous Knowledge Processing Lab, 2024, "Sentence transformers," sbert.net
- van den Broek, E., A. Sergeeva, and M. Huysman, 2021, "When the machine meets the expert: an ethnography of developing AI for hiring," *MIS Quarterly* 45:3, 1557-1580
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, 2017, "Attention is all you need," *Advances in Neural Information Processing Systems* 30
- Zandi, S., K. Korangi, M. Óskarsdóttir, C. Mues, and C. Bravo, 2024, "Attention-based dynamic multilayer graph neural networks for loan default prediction," *European Journal of Operational Research*, September, <https://tinyurl.com/yc3p66df>
- Zhang, D., Y. Yu, J. Dong, C. Li, D. Su, C. Chu, and D. Yu, 2024, "MM-LLMs: recent advances in multimodal large language models," in Ku, L.-W., A. Martins, and V. Srikumar (eds.), Findings of the Association for Computational Linguistics ACL 2024, Association for Computational Linguistics
- Zhang, J., Y. Xie, W. Ding, and Z. Wang, 2023, "Cross on cross attention: deep fusion transformer for image captioning," *IEEE Transactions on Circuits and Systems for Video Technology* 33:8, 4257-4268

© 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 management and technology consultancy specializing in driving transformation in the energy and financial services industries. 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 LinkedIn, Instagram, Facebook, and YouTube.

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
Houston
New York
Orlando
Toronto

SOUTH AMERICA

São Paulo

THIS UNIQUE IMAGE WAS GENERATED USING MID-JOURNEY, STABLE DIFFUSION AND ADOBE FIREFLY

WWW.CAPCO.COM



CAPCO
a wipro company