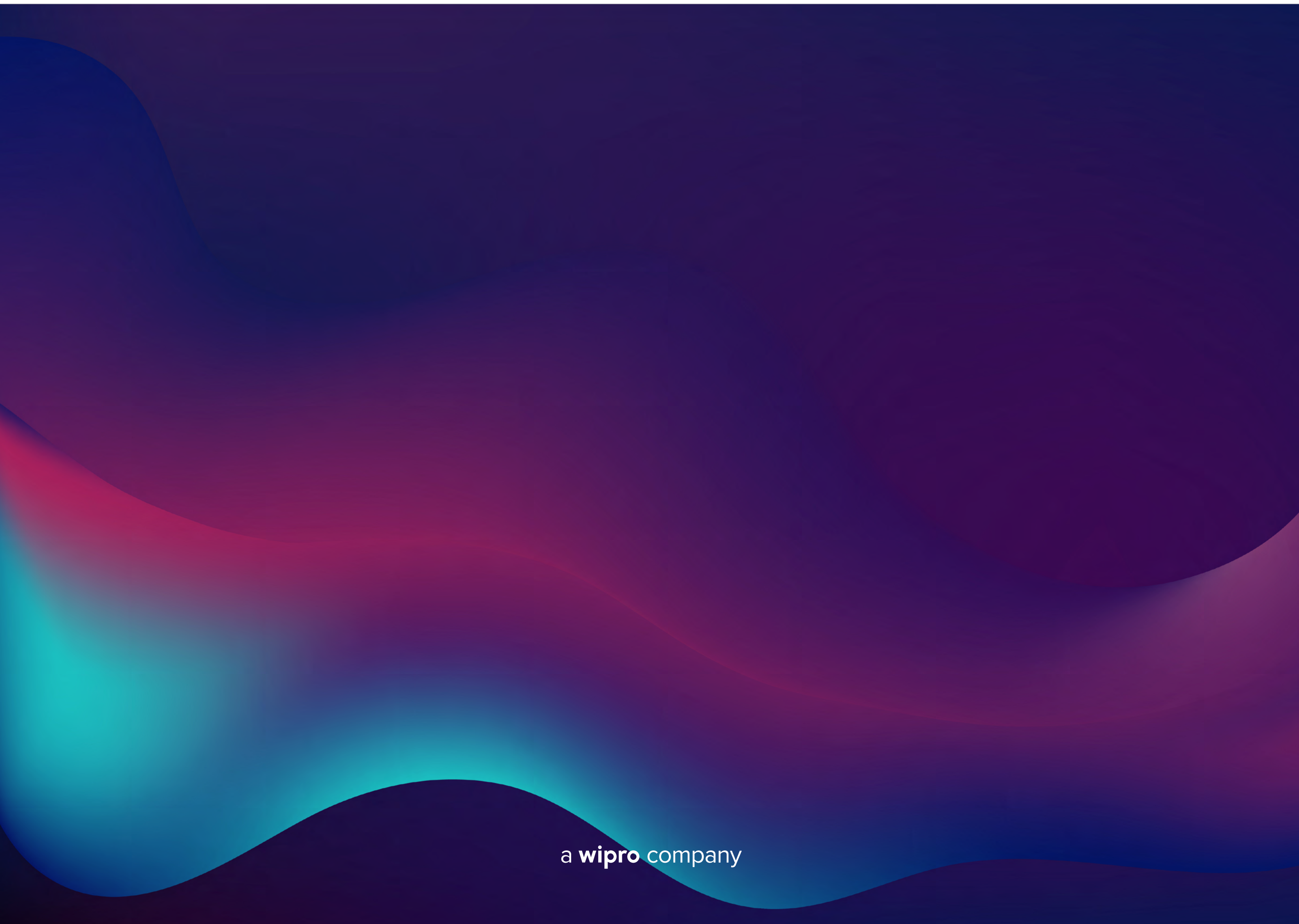
CAPCO

AI GOVERNANCE IN FINANCIAL SERVICES AFTER THE EU AI ACT



a wipro company

In November 2023, we published a paper on applied AI governance that demonstrated that a data-driven perspective in combination with newly introduced roles around AI governance and AI compliance can largely de-risk the implementation of AI solutions.[1]

Following the adoption of the EU Artificial Intelligence Act in March, we have taken the opportunity to reappraise our AI governance framework to validate that it meets these new regulatory requirements.[2] In this new paper, we show that this type of strict AI governance approach can provide the necessary control framework to comply with the new Act and serve as a checklist for ensuring ongoing compliance.

Intended to provide a common regulatory and legal framework for artificial intelligence, the Act aims to establish trustworthy AI while fostering a safe and innovation-friendly environment for users, developers, and deployers. It applies to AI usage within the EU, regardless of the provider's location, similar to the principle established by GDPR.

The Act will come into force 20 days after being published in the Official Journal of the EU, expected to take place in June 2024.

Implementation will occur gradually, with higher risk categories prioritized in the Act's roadmap (see overview in the table below; detailed descriptions of categories and their application in financial services provided in the Appendix).

AI risk category	EU AI Act timelines
Unacceptable	Six months after entry into force, AI in this risk category must be banned.
High	24 – 36 months after entry into force, AI in risk category must comply with regulation.
General purpose AI	12 months after entry into force, AI in this category must comply with regulation. Additional regulation applies by the end of 2030 for certain AI systems that are components of large-scale IT systems.
Limited risk	Transparency obligations are already covered under current regulations.
Minimal risk	No regulation applies.

The EU Commission emphasizes the benefits of integrating regulation as a baseline for development of AI applications. It mandates competent authorities, as defined in various EU financial services directives, to supervise and enforce the regulation.

These authorities are tasked with ensuring compliance and market surveillance of AI systems used by regulated financial institutions, with the option for member states to assign this responsibility to other designated bodies. The authorities are granted powers under relevant regulations to conduct surveillance activities and enforce compliance effectively.

The EU legislators clearly imply that financial services already fall under extremely strict regulations and supervision. The EU AI Act does not impose new obligations or changes to the general regulatory framework (e.g. DORA, MiFIR/MiFID, EBA GLOM) – but rather provides an additional regulation which is supervised in a similar way to previous regulations.

GOVERNANCE PRIORITIES

The EU AI Act explicitly states many examples of the ways governance applies to AI – for instance, in ensuring that AI systems are built in a compliant way, that training data is of good quality technically (and does not raise ethical, societal or environmental concerns), that the AI systems' output is explainable, traceable, and respects IP issues, and that humans interacting with AI are always aware of this fact.

In addition, governance requires that all these aspects are continuously tracked, and if breaches or unforeseen scenarios materialize, then the deployers of such technologies must report such occurrences to the specified authorities.

The EU Commission further repeatedly stresses the importance of privacy, transparency, diversity, non-discrimination, fairness, and social/environmental wellbeing in AI development and usage. These standards demand compliance with existing privacy laws, with regard to the transparent use of AI, the avoidance of discriminatory impacts, and the consideration of societal and environmental impacts.

These principles serve as guiding principles during the design and usage of AI models whenever feasible. In addition, the Act states that "high quality training, validation and testing datasets require the implementation of appropriate data governance and management practices. Training, validation, and testing datasets should be sufficiently relevant, representative, and free of errors and complete in view of the intended purpose of the system."[4]

Lastly, the EU Parliament has highlighted specific concerns regarding generative AI (GenAI) in its requirements for the AI Act. GenAI, which includes large language models (LLMs), is rapidly proliferating, leading to potential misuse of datasets and intellectual property concerns. There is a risk of copyright infringement through outputs generated by GenAI or deep fake technology, with the latter posing significant threats by deliberately creating deceptive content, contributing to disinformation, and misleading audiences. The EU Parliament referred to the need for governance in this context during the consultation stage of the regulation.[5,6]

CAPCO'S AI GOVERNANCE MODEL

Our AI governance model fulfills the requirements of the new EU AI Act, including the use of AI itself within AI governance to achieve compliance in an effective way.

Automation is key

Our AI governance model is a good practical starting point for financial services institutions setting out on their compliance journey. The three key outcomes that any AI governance model needs to deliver to comply with the EU AI Act are as follows:

- Explainability of AI results (where explainability provides transparency, allows bias detection and bias mitigation, supports error detection and error correction, and establishes ownership of intellectual property rights in AI-generated content)
- Protection of intellectual property and sensitive data (including protecting confidentiality of sensitive information and compliance with other regulations, e.g. GDPR)
- Combatting hallucinations.

Our solution can be compartmentalized into three sub-solutions that directly correlate to the three key outcomes. The key lies in automation, which requires a conceptual approach and enables standardized, repeatable, and impartial procedures to ensure results are measurable and comparable.

Automation can be successfully used in respect of all three above-mentioned outcomes for AI governance, specifically:

- Creating a trace between AI outputs and input materials via metadata and contextual information
- Protecting intellectual property through end-to-end automation in information processing
- Combating hallucinations by incorporating external knowledge, data augmentation, and smart prompting (automation in this context is simply achieved by using available technologies such as retrieval-augmented generation (RAG), or the provisioning of additional semantics by generating and deploying knowledge graphs).

The figure below provides a high-level overview of our AI governance model (detailed descriptions of processes involved can be found in our original paper).[1]

Although automation alone will not solve all challenges associated with an AI governance model, it will have a crucial impact on reducing repetitive, routine tasks and freeing resources to focus on more complex tasks.

		PROCESS: INFORMATION / KNOWLEDGE LIFECYCLE						
		01 Pre-processing	02 In-processing					03 Post-processing
SOLUTION OVERVIEW – GOVERNANCE ACCELERATOR		Raw input information is tagged/annotated	CI/CD Pipeline for knowledge to automatically process input	Graph database (reflects the knowledge domain)	Export the database to structured text format	Python script turns export into natural language	Custom training ("finetuning") or API for prompt engineering	Guardrails are applied to Gen-AI
GOVERNANCE TRACE	OUTCOME	a. Being able to explain the results that AI provided	b. Protecting intellectual property and sensitive information					c. Combating hallucination
	CONTROL	[a1] There is a clear trace between input (information given to AI) and the result returned by AI.	[b3] The Gen-AI model and information processing pipeline are deployed in an environment with restricted access.			[b2] Data anonymization and redaction techniques are employed to remove any identifying information from the input data.		[c1] Guardrails are applied to the models to prevent it from providing information outside given boundaries.

AI governance automation solution overview[1]

ROLES AND EXECUTIVE PARTICIPATION

Another important ingredient to successfully deploying AI is the introduction of new roles and project members. As regulatory requirements affect the development and usage of AI in general, and generative AI in particular, new roles will need to be established to achieve regulatory compliance. This is particularly relevant for the financial services sector.

The necessity of making governance-relevant tasks an integral part of software development lifecycle (SDLC) projects is ever increasing, and particularly so since the emergence of GenAI and AI regulations. New roles that would be essential to support these processes include:

- **AI Governance Lead:** Oversees the implementation and execution of AI governance models, including continuous monitoring and quality of tracking and reporting.

- **AI Risk Manager:** Identifies and assesses risks associated with AI, quantifying, mitigating, tracking, and reporting them.

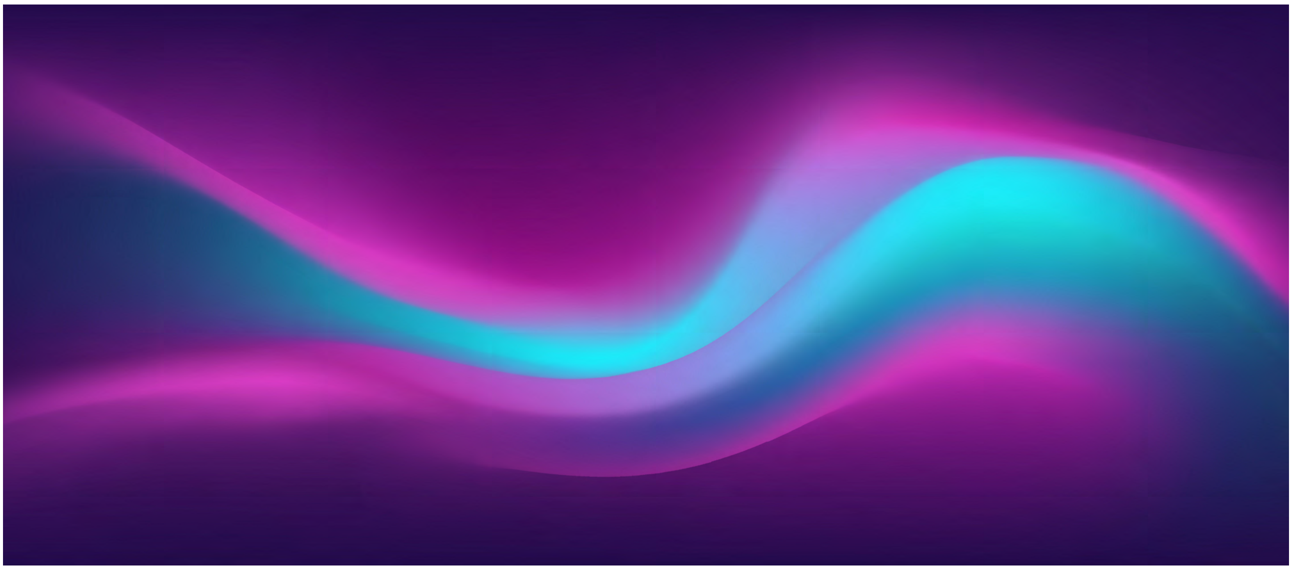
- **AI Compliance Officer:** Ensures compliance with relevant laws and regulations, including reporting obligations to authorities (entering AI-systems and their metadata into EU databases, and incident reporting).

- **AI Ethics Officer:** Ensures ethical use of AI, including monitoring the usage of ethical, sustainable data, and assuring ethical output.

- **AI Technical Architect:** Designs and implements technical infrastructure supporting the AI governance model.

While some of these roles overlap, they require vastly different expertise and project organization, not all of which are yet common knowledge or best practice.

Capco has developed an approach to integrate these roles within financial organizations in a practical way. As the implementation and development of AI and relevant applications steadily increases, the above roles and responsibilities may evolve (i.e. staffing and responsibilities can overlap, and a single person or unit can cover multiple roles), however, fundamental responsibilities stemming from the governance framework will still need to be addressed.



CONCLUSION

As AI continues its rapid evolution, regulations can be expected to keep pace. Article 73 of the EU AI Act empowers the EU Commission to amend key aspects of the Act, including the definition of high-risk AI systems, technical documentation requirements, conformity assessment, and the EU declaration of conformity.

Continuous quality and risk management by providers of products or services incorporating AI are essential for post-market releases to uphold trustworthiness and compliance. While currently most AI systems in financial institutions are deemed low-risk, firms must be cognizant that future developments could alter this classification.

Newly created governance roles – as well as existing compliance departments – can be assisted in their compliance efforts by artificial intelligence itself. For example, GenAI techniques can be used to monitor and analyze the output of regulatory bodies, indicating which parts of new or updated regulations become applicable. Furthermore, the relevant application, process or product owners can be notified, and compliance measures identified and planned accordingly.

By enhancing the structure of their compliance setup to ensure a robust governance framework, financial institutions can be confident of navigating and absorbing the impacts of future regulatory change.

HOW CAPCO CAN HELP

Capco's approach of implementing new governance roles and increasing the degree of automation fulfills the regulatory challenges of the new EU AI Act and aligns with the regulation's forward-looking nature.

Our framework uniformly covers compliance with existing regulations' requirements, such as MaRisk, BAIT, DORA and MiFID. These requirements have not as yet been superseded by the EU AI Act. With a proper governance structure as an integral part of your AI or GenAI projects setup, you will be prepared for future regulatory initiatives.

Contact us to find out how Capco can get you started in this process, from the definition of an AI governance framework and new governance roles to automated monitoring and implementing regulatory change.

APPENDIX 1/2

EU AI Act risk categories classification and timelines

AI risk category	Description	Purpose	Regulation requirements	Examples in financial services	Timelines
Unacceptable	AI applications engaging in prohibited activities such as manipulating human behavior, conducting real-time remote biometric identification (e.g. facial recognition) in public areas, implementing social scoring systems, etc.	In general, the purpose of the EU AI Act is to protect elementary human rights and to prohibit the usage of AI to violate these.	The EU AI Act explicitly prohibits the usage of AI in this risk category.	Using computer vision or biometric recognition for credit worthiness scoring is forbidden under the EU AI Act.	Six months after entry into force, AI falling into this risk classification must be banned.
High	AI applications posing significant threats to health, safety, or fundamental rights, especially in sectors like health, education, recruitment, critical infrastructure management, and law enforcement.	AI systems originally designed to decrease threats, ensure critical infrastructure, surveil health, education, or law enforcement can easily backfire when malfunctioning or abused. In such cases they may achieve the opposite: putting humans at risk, cutting access to health, education, or critical infrastructure, or –in the context of law enforcement - putting everybody under general suspicion of being a criminal (reversing the maxim that everybody is innocent until proven guilty).	Requirements include adherence to quality and security standards, human oversight, and constant evaluation, starting before the placement on the market and throughout the lifecycle and incident reporting. Also, the list of high-risk applications and systems may be expanded without amending the AI Act itself. According to Article 73, providers of high-risk AI systems placed on the EU market must report any serious incident to the market surveillance authorities of the Member States where that incident occurred.	Fraud detection and anti money laundering will put everybody under suspicion and in case of malfunctioning (false positives) may restrict access to money or put humans into the position of having to prove their innocence. AI based creditworthiness assigns scores to repayment probabilities which may result in systematic and unethical exclusion of groups of people. Related questions turn up in the context of using AI to calculate health and life insurance premiums. Consequently, any AI system used in a bank's credit and loan operations, or anti money laundry operations falls under this category. Therefore, banks' compliance departments must establish robust AI governance specifically for such activities to meet regulatory requirements.	24 months after entry into force, AI falling into this risk category and as described in Annex III (e.g. biometrics, critical infrastructure, administration of justice and democratic processes educational training, etc.) must comply with regulation. 36 months after entry into force, AI falling into this risk category and as described in Annex II (list of criminal offences) must comply with regulation.

APPENDIX 2/2

EU AI Act risk categories classification and timelines

AI risk category	Description	Purpose	Regulation requirements	Examples in financial services	Timelines
General purpose AI	This category has been added in 2023 in view of emerging GenAI technologies that became publicly available (ChatGPT). It does not typically pose threats in the sense of high risk category, however, it produces content in a persuasive way. This category can be considered synonymous with generative AI which generates output from vast amounts of input data. General-purpose AI poses systemic risks when the cumulative amount of computation used for its training exceeds 10^25 FLOPS.	For users directly or indirectly interacting with this AI, different input/output modalities are possible, e.g. text-to-text, text-to-image, multi-modality. The output is generated in a dialogue form by answering questions or performing tasks. The EU will provide sandboxes for checking such applications before placing them on the market.	Providers of general-purpose AI must provide comprehensive technical documentation, a policy to comply with the EU copyright law, and a sufficiently detailed summary of the content used for training. Providers of general-purpose AI with systemic risks must furthermore perform model evaluation and testing to mitigate systemic risks.	Typical use cases from the financial sector include personalized communications, KYC risk assessment, policy drafting, regulatory compliance, and knowledge base assessment. Risks derive from missing transparency, bias, hallucinations, or IP infringement.	12 months after entry into force, AI in the category must comply with relevant regulation. By the end of 2030, obligations go into force for certain AI systems that are components of large-scale IT systems, established by the EU law in the areas of freedom, security, and justice, such as the Schengen Information System (Article 83).
Limited risk	Such AI-applications do not pose immediate threats to health, infrastructure, etc., as they typically inform humans on certain topics (e.g. Q&A chatbots) or serve for entertainment (audio and video).	This category gives specific exceptions to the high-risk category by describing scenarios where AI is supplementing human work, for example narrow procedural tasks (converting unstructured data into structured data, pre-classification tasks, or de-duplication), preparing content that will be used by a human (translation, transcription), polishing human-generated content (spell-checking, formatting). This category also contains chatbots, AI-generated texts for information on matters of public interest or entertainment, as well as audio and video content that are known as deepfakes.	The AI Act introduces specific transparency obligations to ensure that clients are informed about the usage.	Typical use cases within the financial sector include biometric verification, translations, pre-classifications, chatbot functionalities and message assembling systems.	For financial institutions, transparency obligations are already covered under current regulations.
Minimal risk	Everything not covered elsewhere falls under this risk category.	AI-enabled games, spam-filters	No regulation	Examples include internal monitoring tools for spam or phishing mails.	No regulation applies

5. REFERENCES

1. https://www.capco.com/Capco-Institute/Journal-58-Artificial-Intelligence/Applied-Generative-AI-Governance

2. ibid.

3. https://artificialintelligenceact.eu/

4. TA (europa.eu)

5. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf

6. Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity by Claudio Novelli, Federico Casolari, Philipp Hacker, Giorgio Spedicato, Luciano Floridi :: SSRN

AUTHORS

Peter Lehnen, Principal Consultant
Igor Schnakenburg, Managing Principal
Gerhardt Scriven, Executive Director
Alessandro Corsi, Partner

CONTACT

Alessandro Corsi, Partner
M +49 174 695 4839
E Alessandro.corsi@capco.com

Gerhardt Scriven, Executive Director
M +55 11 99698 5649
E Gerhard.scriven@capco.com

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