Finance Watch

Making finance serve society

Artificial intelligence in finance: how to trust a black box?

Can and should the provision of AI-powered financial services be regulated?

A Finance Watch Report

March 2025

Author: Thierry Philipponnat
Editor: Robert Nosker
Graphics and typeset: Camila Dubois
Cover photo: Adobe Stock

Acknowledgement: We extend our gratitude to the members of Finance Watch for their invaluable input and feedback, as well as to the numerous professionals and experts who contributed to the production of this report by sharing their insights and expertise.

Finance Watch has received funding from the European Union to implement its work programme. There is no implied endorsement by the EU or the European Commission of Finance Watch's work, which remains the sole responsibility of Finance Watch.

Co-funded by the European Union

Contents

Preface - Report objective

Artificial intelligence (AI) is everywhere. Use cases of artificial intelligence in the financial services sector are mushrooming.

This report looks into the compatibility of this emerging practice with existing financial regulation.

Is AI a more efficient way or a new way of conducting financial business?

Can the same rules apply with the development of AI-powered financial services?

Are we at a moment in time where, under the disguise of a new technology, we are effectively confronted with a different way of providing financial services necessitating a rethink of existing regulation?

This report asks three questions: Is AI a new way of conducting financial business, or is it "only" a quantum leap in computing power driving efficiency gains for financial services providers? Is reopening financial regulation in recognition of the evolutions brought to financial services by AI desirable? Are AI-powered financial services compatible with financial regulation in general and with the EU rule book in particular, and if yes, how?

In the following pages this report will look into the functioning of AI, analyse the consequences of its modus operandi when applied to financial services and activities, and look into the desirability and the possibility of regulating AI-powered financial services. This can be considered as the first step of the gap analysis to be conducted if EU financial regulation is to be reopened.

Important note: this report was human-produced apart 1) from two definitions provided in the glossary (in which case, the large language model (LLM) used is indicated as a source), and 2) from Annex I, which is a non-edited copy / paste of responses produced by LLMs on the subject of this report.

Executive Summary

The adoption of AI in the financial services sector is on the rise. While AI itself is merely an umbrella for a variety of algorithm-based technologies that solve complex tasks by carrying out functions that previously required human thinking, its adoption in the financial sector promises enormous efficiency gains to financial service providers (and hence, more profits). The public interest dimension of AI is hardly evident.

This report examines the challenges and vulnerabilities that arise with the adoption of AI in financial services. It additionally makes policy proposals to address these risks, ensuring the protection of consumer interests and safeguarding the stability of the financial system as a public good.

The very nature of AI is at odds with the fundamental principles underlying decision-making in finance and financial regulation itself – accountability, responsibility and transparency. First, there is a fundamental difference between human intelligence and the logic of operation of technologies which came to be called "artificial intelligence". This difference essentially boils down to the fact that AI draws conclusions and makes predictions based on the correlations detected in the data used for AI model training (inductive logic), whereas human brains are able to reach conclusions with a much broader set of methods, including deduction, abduction, as well as heuristics or intuition.

Second, the current system of financial regulation (and consequently the financial system itself) is built on the principle of causation, i.e. establishing a logical relationship between the cause and the consequence in every individual case (and not in the aggregate, which is the correlation logic used by AI). The causation principle itself is a pre-condition to a rule-based system with accountability, responsibility and transparency.

By contrast, AI technologies operate as "black boxes" – generating outputs without clear explanations of their reasoning. By definition, such outputs are not accessible to human analytical ability. This renders the ability of humans to intervene impractical, if not impossible, and is particularly problematic. If AI-driven credit assessments, insurance pricing, or investment products cannot be adequately explained, customers, investors, and regulators may struggle to detect errors, biases, or systemic risks. Without robust explainability requirements, AI-powered decisions could undermine confidence in the fairness of financial markets.

Data manipulation and the inevitable exhaustion of human-generated data to train AI models are further challenges, which present themselves in the context of AI adoption, leading to nonsensical results and the un-detectability of falsehood.

Questions arise as to whether AI operates beyond the reach of existing legal frameworks. In order to address this challenge, the report proposes key principles

for regulating AI-powered activities and products: i) either prohibiting the activity or service where risk is unacceptable from the public interest perspective or ii) regulating the provision of service / activity at data and data governance level, i.e. establishing rules around the process of development and deployment of AI.

The EU AI Act, applicable from August 2026, is largely based on the abovementioned principle. However, financial activities are only very partially covered by this legislation; thus its scope needs to be extended to all financial services for the Act to become meaningful for the financial sector.

This report presents an analysis of vulnerabilities created by AI. This should serve as a basis for a detailed gap analysis of the existing EU legislation to determine which regulations must be reopened and which amendments must be made to ensure that the interests of investors, consumers, citizens and society at large are protected in a world of AI-powered financial services. Such vulnerabilities range from threats to consumer protection, risks to financial stability, and complexities in supervision and fraud detection.

In the area of retail finance, deployment of AI could lead to opaque credit assessments, pricing discrimination, discriminatory lending, and misleading financial advice, resulting in financial exclusion that disproportionately affects vulnerable consumers. Moreover, financial institutions are increasingly dependent on third-party AI providers and face operational risks from unregulated external systems, as well as concentration risks stemming from a small number of dominant AI firms' control over critical models and infrastructure. Deployment of similar algorithms by financial institutions further aggravates the risk of herding in financial markets. Using AI models as risk management tools reinforces the ill-founded sense of security given to (bad) risk-managers by models. Finally, supervisors face a challenge of being able to keep pace with the deployment of AI by financial institutions and deliver on their mandates.

Without clear regulatory guardrails and accountability mechanisms, the usage of AI in financial services introduces risks that are difficult to detect and control, threatening consumer protection and market stability while undermining trust in the wider financial system. AI is here to stay. Thus the focus must now be on prevention, not damage control. The public interest cannot be sacrificed for the sake of private sector profitability or arguments of competition. To lessen the risks of AI-powered financial services, this report makes recommendations to navigate the trade-off between maximising AI's efficiency gains and ensuring broader financial and societal safeguards.

Key Recommendations

Adopted in 2024 to address the challenges of AI, the EU AI Act – the world's first comprehensive AI law – represents a groundbreaking legislative effort to regulate artificial intelligence. While a step in the right direction, financial activities are only partially covered in the AI Act. Thus, Finance Watch calls upon European policymakers to:

1

Broaden the EU AI Act

The European Commission should extend the scope of high-risk AI systems in the AI Act:

→ Broaden the scope of Annex III of the AI Act to cover all financial services

→ Adopt delegated acts as per Article 7 of the AI Act to this effect

In addition, we call on European policymakers to:

2

Establish an AI Civil Liability Regime

The EU should establish non-contractual civil liability rules for AI and reintroduce the AI Liability Directive proposal made by the European Commission in 2022 and withdrawn in February 2025:

→ Align the scope of the liability regime with the revised scope of the EU AI Act defining all financial activities as high-risk (see recommendation n°1)

→ Cover non-contractual civil liability for damages caused by the output of an AI system or by the failure of an AI system to produce an output

→ Include a rebuttable 'presumption of causality' reversing the burden of proof and making legal redress technically possible for claimants

3

Evaluate the potential for supervisory enforcement

→ The European Commission and EU supervisors should assess legally and technically the possibility for financial supervisors to enforce existing EU financial regulation for AI-powered financial services.

4 **Conduct a regulatory gap analysis**

→ The European Commission should conduct a gap analysis to determine which financial regulations must be reopened to ensure that the interests of investors, consumers, citizens and society at large are protected in a world of AI-powered financial services.

→ EU policymakers should amend the relevant legislative texts according to the results of the gap analysis conducted.

Introduction

Revolutionary! New paradigm! Game changing! Pioneering a new era! Innovation will save the world! Efficiency! Competitiveness! Productivity! Operational efficiency! Client engagement! Sustainable growth! Increased efficiency! Cost savings! Improved risk management! Enhanced revenue generation! Disruptive! New frontier!

Reading the numerous artificial intelligence-related publications from business, consultants, academics and international organisations, there seems to be little doubt: artificial intelligence (AI) is about the eternal race towards ever-greater efficiency and more profits, in other words about business.

In its AI in Finance 2024 GPS report,[1] Citi asserts that "*AI could add $170 billion or 9% to global banks sector profit pool by 2028*". The motivation is clear.

Is there any room for public interest in the exponentially expanding world of AI-powered financial services? What's in it for the public interest? Are the private business interests of both AI solutions providers and financial institutions on one side, and the interest of consumers and citizens on the other side compatible? Do we need to adapt the rules to ensure they are?

Invariably, the papers published on AI in financial services describe the efficiency gains to be expected, only to speak about the risks secondarily. Interestingly, AI is always presented as "good for business despite the risks" but, beyond lip service encountered here and there, never convincingly as "good for customers, for citizens, for the public interest, for financial stability, for the planet, etc.". Could it be that, when it comes to artificial intelligence in financial services, public interest would be at best in damage control mode?

This report starts with the observation that AI is not only about computers developing unprecedented levels of processing power but, even more importantly, about a different approach to treating data and deriving conclusions from the data processed.

If the use of AI in financial services was solely about injecting processing power far beyond the power of the human brain into financial activities, fewer questions would arise. The reality is that AI can change, and to some extent already has changed, the way the financial industry operates.

If the objective had been "how to promote the public interest in finance", the answer would not have been "artificial intelligence". Having said that, going back is not a credible option: AI is already part of financial services practices and it will increasingly be so in the future. The question now is whether we should conceive a financial regulation and a financial supervision able to protect the public interest in this new AI en-

1 Citi GPS, *AI in Finance – Bot, Bank & Beyond* – June 2024.

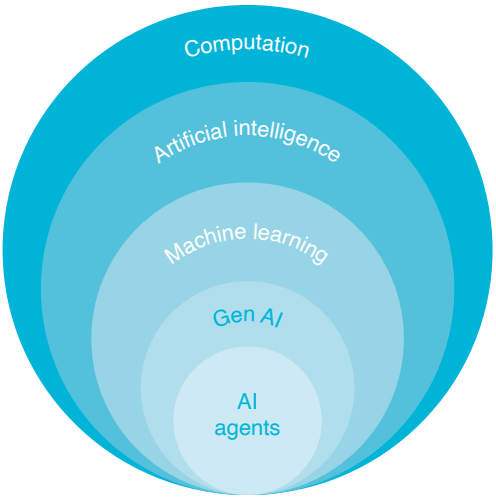
vironment. And, if yes, how. How does society move beyond damage control mode? How can regulators and supervisors go beyond applying a plaster on a wooden leg? How can policy-makers build an approach going beyond an inherently weak defensive position struggling to catch-up with innovation?

I. Understanding artificial intelligence and apprehending the difference with human intelligence

A. What's in the two words artificial intelligence? A brief description and a glossary of AI

AI's landscape can be summarised by Figure 1[2] below:

Figure 1: Decoding AI



Definitions[3]:

AI agent: as a system or program capable of perceiving its environment and performing tasks in an autonomous manner, an AI agent could be considered as the ultimate stage of artificial intelligence.

Artificial intelligence (AI): umbrella term for a range of algorithm-based technologies that solve complex tasks by carrying out functions that previously required human thinking (UK Information Commissioner's Office).

Deep learning: a subset of machine learning that focuses on using artificial neural networks with multiple layers (deep networks) to model and learn from complex data representations (ChatGPT).

Foundation models: models trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks (Stanford).

---

2 Source: BIS Working Papers, No 1194 – *Intelligent financial system: how AI is transforming finance*, June 2024. Figure 1 page 5.

3 These definitions are largely, if not exclusively, inspired from the "What are we talking about when we talk about AI?" section of the Open Markets Institute report '*AI in the Public Interest: Confronting the Monopoly Threat*' of November 2023.

General purpose AI: largely synonymous with foundation models; refers to an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed (European Parliament).

Generative AI (GenAI): describes algorithms and applications that can be used to create new content, including audio, code, images, text, simulations, and videos (McKinsey).

Large language model (LLM): AI systems trained on significant amounts of text data that can generate natural language responses to a wide range of inputs (Ada Lovelace Foundation). Preserving the ability of LLMs to model low probability events is essential to the accuracy (and for societal issues, the fairness) of their output. Unfortunately, when LLMs are trained on "synthetic" (i.e. derived from AI) datasets, that ability vanishes. Training on samples generated from another generative model can (and does) induce a distribution shift, also dubbed 'loss of variance', where the majority of sub-populations become over-represented at the expense of minority groups.

Machine learning (ML): a subset of AI that focuses on developing algorithms and statistical models that enable systems to learn from and make decisions or predictions based on data, without being explicitly programmed for specific tasks (ChatGPT). The algorithm can be trained on structured data (supervised learning) or unstructured data (unsupervised learning). ML is about learning from examples rather than instruction. Technically, it is about creating a model whose average statistical error is the lowest possible.

Key technological developments in AI[4]

Early uses of AI. AI and ML have been used in the financial sector long before the advent of LLMs and GenAI. These earlier technologies were primarily used for automating routine tasks, detecting fraud, and making predictions based on historical data. ML, a subset of AI, employs algorithms to learn from data and make decisions or predictions. Neural networks, a technology used in a specific type of ML and inspired by the human brain, are particularly effective at recognising patterns and complex relationships in large datasets. Deep learning, a further subset of ML, uses multi-layered neural networks to learn and extract complex patterns from large data sets, thereby significantly enhancing the accuracy of predictions. These foundational technologies paved the way for more advanced applications like LLMs and GenAI.

4 Source: Box 1, page 5 of the Financial Stability Board report *The Financial Stability Implications of Artificial Intelligence*, November 2024.

Advent of LLMs and GenAI. The launch of consumer-facing advanced AI systems like LLM chatbots in November 2022 demonstrates how rapidly the field of AI can experience significant technological change. LLM chatbots are a specialised application of generative AI with focus on language, whereas GenAI models are able to generate new content, such as images, text or video, based on user prompts.

Role of Natural Language Processing. LLMs are an advanced application in the broader field of natural language processing (NLP), which is concerned with enabling machines to recognise, process, and understand the content and meaning of language. NLP has been used by firms, including those in financial services, for many years for customer interaction, regulatory compliance monitoring, automated advice, and sentiment analysis on customer feedback. Prior to LLMs, the most advanced forms of NLP operated by transforming words into individual units, or "tokens", and then translating the tokens into numerical representations (vectors) that attempt to capture the meaning of words, a process known as "word embedding". However, there are limitations to this approach: words can have multiple meanings (e.g. "interest" can refer to "attentiveness" or "interest rates") and the context of an entire text can provide more nuance around the meaning of a word compared to focusing on surrounding words.

Features of the Transformer. The Transformer, a deep learning architecture and one of the foundational technologies of LLMs, addresses the aforementioned limitation of word embeddings by incorporating an attention mechanism, which focuses the neural network on specific parts of the text. For example, it can differentiate the meaning of the word "bark" depending on whether it is used in a sentence about dogs (e.g. "the dog has a loud bark") or trees (e.g. "the tree's bark is brown"). Positional encoding is used to understand the order of words in a sentence. LLMs also have other features such as the feed-forward mechanism, which helps to predict the next word based on previous words, and a system that identifies the most likely next word occurring in a sentence. Altogether, these components, particularly the attention mechanism and positional encoding, enable LLMs to process and generate text more efficiently. The Transformer also processes data in parallel, rather than sequentially, which contributes to further efficiency gains compared to older models.

Limitations and future prospects. While LLMs and GenAI are able to mimic human language and creativity, the currently available models do not truly understand the content they generate. This is because their outputs are the result of a stochastic process rather than a deep understanding of the underlying text. The field of AI is dynamic and future advancements, potentially emerging from other AI sub-fields, could reshape the landscape and impact the financial system in ways that are not fully predictable at present. These future developments could introduce new vulnerabilities and challenges for financial stability, underscoring the importance of continued monitoring, research, and policy consideration by financial authorities.

B. The difference between human and artificial intelligence: the nature of AI

Human intelligence is the ability human beings have to understand, develop abstract reasoning, think critically, solve problems, make appropriate decisions, adapt to a changing environment, and deal with new and unexpected situations, including situations that have never been seen before.

"Intelligence" comes from the Latin "intelligentia" deriving from "inter" ("between") and "legere" (to pick out, to choose, to read). Being intelligent is being able to "choose between". This not only a Western conception; witness the Japanese word "wakaru" (分) meaning both separating and understanding.

Human intelligence operates along three types of inference: deduction, induction and abduction. Deduction leads to a conclusion necessarily true as it derives logically from its premises. Induction is a generalisation process whereby a conclusion is drawn from a set of observations. Abduction derives a likely conclusion from an observation.

Despite the usage of the word intelligence, artificial intelligence is altogether a completely different process from human intelligence.

Artificial intelligence operates, by and large, on induction whereas human intelligence uses in an indiscriminate manner deduction, induction and abduction. AI may, in some circumstances, use abduction to derive plausible or probable explanations as a complement to inductive reasoning, and it uses deductive reasoning in logic programming (so-called Prolog), but those instances remain marginal.

Inductive reasoning is fundamental to how AI models generalise patterns, make predictions, and derive conclusions from the data they are trained on. Induction is at the core of machine learning and deep learning, but also of pattern recognition, predictive analytics, neural networks and transfer learning.

Machine learning and deep learning detect patterns and draw correlations from the data the models are trained on and produce results based on **inductive reasoning:**

the patterns and the correlations detected on the data used for model training are extrapolated to produce the results. Consequently, large language models (LLM) operate on an inductive logic. The fact that the amount of data processed to derive the results is gigantic does not change the nature of the process: LLMs infer general conclusions from the dataset they analyse and this is what induction is about.

Importantly, inductive reasoning itself is effectively a different process in an AI or in a human context. AI models perform inductive reasoning in a statistical and data-driven way, whereas humans use more **intuitive or heuristic methods.** Heuristics, the art of making discoveries, solving problems or apprehending the world with a limited and knowingly incomplete dataset, is particularly important for human reasoning. It is in sharp contrast with artificial intelligence and its necessity to treat enormous amounts of data to derive results.

AI needs a large amount of data to generalise effectively, whereas humans can generalise from a limited set of examples (data) and, in extreme cases, without data (a human being is able to deal with entirely novel situations). This not only makes artificial intelligence not intelligent in the human sense of the word, but it also makes AI models' quality of output only as good as the quantity and the quality of the data they were trained on.

Data processing is not only what makes AI inherently different from human intelligence, it is also its weak spot as AI cannot operate without large quantities of data and **flawed data is synonymous with flawed output.** Incidentally, AI models' outputs rely on the assumption that the future will look like the past, which can be a reasonable assumption in some fields but is widely recognised as a bold one in economic and financial matters.

The history of philosophy can help us understand one or two things about the difference between human and artificial intelligence. This will prove useful when analysing the consequences of using AI in the field of financial services and its impact on financial regulation.

Some of the most elaborate philosophical endeavours in the history of mankind, notably if far from exclusively, those by Descartes and Husserl, have been articulated around the idea of thinking without prior knowledge nor experience and "suspending judgment" (epoche); in modern speak, we would say without data. For those philosophers, human intelligence is not about regurgitating previously absorbed information or inferring from it but precisely the opposite: what conclusions can human intelligence draw without prior knowledge of a phenomenon? Strikingly, this is exactly the opposite of artificial intelligence.

From Sextus Empiricus and Hume to Popper and Russell, to name but a few, philosophers have questioned the rationality of making predictions on observations of the past, in other words the validity of inductive reasoning. What can make us say that the future will reproduce the past? Many philosophers challenge this possibility on the basis of the inevitably limited number of observations used as a reference and on the

weakness of the assumption that the future will be a reflection of the past. The question, and all its subtleties, is also at the heart of Bayesian probabilities. The heart of the arguments developed by those thinkers is that inductive reasoning cannot rationally justify causality. This is often summarised as "the problem of induction".

In the case of AI, and in particular in the case of AI used for financial services, the problem of induction takes two specific dimensions:

1. For AI at large (thus including AI used for financial services), the problem of the inevitably limited number of observations is not so much relevant (the number is enormous), but the problem arises both from the fact that the number of quality (i.e. human-produced) observations is exhaustible, and that the data can be manipulated.
2. In the specific case of AI used for financial services, the weakness of the assumption that the future will be a reflection of the past is crucial. Consumer behaviour and corporate performance can and do evolve for many reasons, and financial market returns are a function of too many different rational and irrational factors to ever be a copy of the past; witness the widely-commented 'fat tail' question of the statistical distribution of returns. With the objective for financial regulation to be 'fair, clear and not misleading', all of this has led financial regulators to require fund documentation for retail investors (UCITS KIID[5]) to mention as a word of caution that past performances are not a prediction of future performances. Strangely, this widely recognised state of the world seems to be often forgotten in the current reflections on the ability of AI to decrypt or predict consumer behaviour, corporate performance or financial markets returns. The fact that detecting patterns and correlations on very large datasets (which are by essence a reflection of the past) does not confer the ability to predict the future seems to be forgotten.

The notion of AI taking over human intelligence and escaping the control of its creator in a Frankenstein-like manner is also a fantasy. For all its enormous processing power, AI is inherently more limited than human intelligence for the reasons already mentioned: AI walks on one leg (induction) whereas human intelligence walks on three (deduction, induction and abduction). Furthermore, AI necessitates a large amount of data whereas human intelligence is able to draw useful and meaningful, if not necessarily scientific,[6] conclusions from a limited set of data. For humans, being overwhelmed by AI would be a choice, but it is not a fatality.

The differences between human and artificial intelligence are, as we can see, very significant. This will have far-reaching consequences when analysing whether AI-generated practices in financial services are compatible with a financial regulation conceived by and written for human intelligence.

AI's biggest marketing trick is, undoubtedly, the choice of its name. The use of the word *intelligence* evokes one of the most defining dimensions of humanity, when it is not.

---

5 Undertakings for Collective Investment in Transferable Securities Key Investor Information Document.

6 An essential reference on this subject is Karl Popper's 'The logic of scientific discovery' published in 1959.

II. Is the provision of AI-powered financial services compatible with a rule book?

AI in financial services is already here and it is only a start. The push towards ever-greater efficiency and productivity is just irresistible for financial services providers. Worse, if a provider made the choice of not going down the AI route, it would risk being left behind by its competitors. Running at least as fast, ideally faster, than competitors is the essence of business. He who does not run is dead. From a finance business standpoint, this is far more important than knowing whether this new way of doing business is compatible or not with the existing rule book or, as a matter of fact, with the very notion of a rule book. Compatible or not with the existing rule book of financial regulation, AI in financial services is here to stay and develop further. **Fait accompli.**

Seeking more efficiency and productivity is not necessarily synonymous with a new way of conducting business. This report argues, however, that in the case of AI-powered financial services, it is. This is due to the inductive nature of AI discussed in Section I. The pre-AI way of conducting financial activity was based on causation, which is a deductive principle. Post-AI financial activity is based on inferring patterns considered to be good enough given the enormous amount of data they are derived from. This is a change of paradigm.

From the public interest perspective, and therefore from a regulator's or a supervisor's standpoint, this raises many questions.

Hitting the epistemological wall of the problem of induction, as described in Section I, is not an esoteric philosophical problem detached from reality. Using AI to provide financial services, manage financial activities or regulate the provision of such services and activities, challenges both the principle of causation implicitly at the heart of financial activity and the possibility for financial regulation of being rule-based. If, as affirmed by the problem of induction, inductive reasoning cannot justify causality, how can the use of AI be justified in finance and regulated by a rule book, both as a matter of principle and from a legal standpoint, given that the architecture of the financial regulatory system is founded on causality?

The question is particularly acute for the principles of **accountability and responsibility**, two founding pillars of financial regulation. Correlation may not be causation conceptually, but in an AI world correlation determines the output, which comes down to saying that causation is no longer the issue. As a consequence, when the financial system starts operating, deriving results or making decisions on correlation, as it does with AI, it operates outside of the causation logic. We seem to have a contradiction at the heart of the system and we need to address it. This will be of particular importance if the development of AI agents and autonomous AI systems over the coming years is as brilliant as many pundits predict.

A. The five challenges of AI

Five idiosyncratic challenges will need to be overcome if we want to make AI-powered financial services compatible with a rule book.

Challenge n°1: the human-in-the-loop paradox.

Often dubbed "human-in-the-loop" or "human-in-the-room", the broadly recognised necessity to inject human intelligence into AI has several dimensions and many different use cases.

Beyond the irony of recognising this necessity given the claim of being artificially intelligent, we can see emerge three dimensions of the human-in-the-loop necessity when AI is applied to financial services:

1. At a general level, human intervention is necessary to obtain the quality data that LLMs can be trained on. This is true with reinforcement learning from human feedback (RLHF), a process part of the fine-tuning of LLMs. This is also true in the broader issue of the foreseeable exhaustion of human-generated data available for the training of LLMs and the subsequent expected model collapse.[7] The paradox of AI is that with its pretension to surpass human intelligence, it feeds from human interaction and collapses when human interaction disappears.
2. At a more granular level of the services actually provided to customers, and in particular financial advice, the human-in-the-loop is supposed to act as a filter to ensure that the financial advice provided – the creditworthiness or the anti-money laundering / KYC (know your customer) assessments made or the pricing of insurance policies – are appropriate for the targeted customer. Beyond its mechanical, negative impact on the productivity of the service provider, this dimension of human intervention in an AI-generated process is both desirable from a customer protection standpoint and technically possible. Noticeably, it has been partially included in the EU Artificial Intelligence Act (EU AI Act)[8] through Article 14 ("Human oversight") in reference to Annex III (High-Risk AI Systems), whose point 5 mentions as high-risk: *"(b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud; and (c) AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance".*
3. When applied to financial activities where AI opens previously unknown horizons or provides otherwise unreachable results thanks to its processing power, the human-in-the-loop logic becomes more problematic, when it is not a sheer contradiction in terms. For instance, using AI to calculate the risk weights of banks'

7 Shumailov, I., Shumaylov, Z., Zhao, Y. et al. *AI models collapse when trained on recursively generated data*. Nature 631, 755–759 (2024).

8 *The EU Artificial Intelligence Act*.

internal models for the purpose of capital requirement calibration under the Basel framework would require the ability to justify the said risk weights to supervisors, which is a contradiction in terms (if the calculation could be done by a human being or traditional techniques, one would not need AI...). Another example could be found in the detection of suspicious or fraudulent transfers of money under the Anti-Money Laundering Directive (AMLD).[9] By construction, a human brain cannot follow the millions of transfers made by a large bank every day (hence the interest of AI). A human can vet suspicious transfers when they are AI-detected, and either confirm their suspicious or fraudulent nature, or requalify them as non-fraudulent / non-suspicious (albeit at the cost of diminished productivity). However, the opposite is not true: faced with a lack of detection of suspicious transfers, a human brain cannot affirmatively state that such transfers should have been detected. The human in the room has found its limit.

In a nutshell, beyond the general affirmation that there is no such thing as AI without human intervention, **human intervention** in AI-powered financial services can be part of the process. As in the case of RLHF, it can be deemed desirable, be technically feasible and be imposed by regulation, such as the human oversight imposed by the EU AI Act for Annex III (high-risk) applications. Alternatively, it may defy the very purpose of AI as in the (far from exhaustive) examples of anti-money laundering and financing of terrorism (AML-FT) detection or Basel framework-related calibration of banks' internal models.

Challenge n°2: is the provision of AI-powered financial services above the law?

From a legal standpoint, human intervention is on paper the solution to the **responsibility and accountability problems** already highlighted. However, the problem arises in the numerous cases where human intervention defies the purpose of AI, i.e. when human intervention is either not technically possible or, if theoretically possible, has the effect of jeopardizing the efficiency gains brought by AI. The odds are that the day when financial institutions have to defend the output of AI activities with their supervisor or in court, they will plead not guilty wherever regulation does not impose human oversight, on the dual argument of the impossibility of human intervention (hence the use of AI) and of the inductive nature of AI processes.

The latter part of the argument can be expected to be of particular importance, given the central role played by causality in supervisors' prosecutions and in court verdicts: it is to be anticipated that we will hear the combined "human intervention was impossible" and **"correlation is not causation"** arguments from the mouths of lawyers defending financial institutions. The absence of causality will lead them to plead that their clients bear neither responsibility nor liability, unless a law establishing liability in such cases is adopted, of course.

9 Directive (EU) 2024/1640 of the European Parliament and of the Council of 31 May 2024 on the mechanisms that Member States should put in place to prevent the use of the financial system for the purpose of money laundering or terrorist financing.

Given that personal responsibility (whether of natural or legal persons) operates on a causation principle (judgments are rendered on the facts and the law), if a person's deeds are not the cause of a damage, whether moral or material, there is a high chance that this person might not be considered as responsible, and holding that person to account might subsequently prove impossible. Absent adequate regulation, this phenomenon can be expected to run in the future in many AI-related cases in opposition to the principles of responsibility and accountability at the heart of our legal and regulatory systems, and therefore against existing legal and regulatory systems.

For instance in the regulatory area, will a supervisor prosecuting a financial services provider in a case of ill-founded financial advice, or failed detection of unlawful transfers of money stand a chance of winning? Will correlation win over causation and thereby win the case for the financial services provider pleading innocent for its ill-founded advice or its transfer of money to recognised terrorists? In that context, **the European Commission and the EU supervisors should assess legally and technically the possibility for financial supervisors to enforce existing EU financial regulation for AI-powered financial services.**

Regarding civil liability, the EU has started addressing AI-related issues with a view of providing individuals harmed by AI technologies with protection and avenues for redress. In October 2024 it adopted an amended version of its 1985 Product Liability Directive (PLD), including software and AI products, but not non-contractual civil liability for damages. In parallel, the proposal for an **'AI Civil Liability Directive' (AILD)** made by the Commission in 2022 had the objective of complementing the PLD to cover non-contractual civil liability for damages caused by an output of an AI system or by the failure to produce an output. Most importantly, the AILD would create a rebuttable 'presumption of causality', which would ease the burden of proof for claimants. Given the black box nature of AI systems, this dimension is essential for the AILD to provide an effective redress mechanism. However, at the time of writing this report the European Commission has announced the withdrawal of the AILD, which otherwise would have been subject to debate and adoption by the EU co-legislators.

Without a liability regime in place, AI technology providers - largely represented by the US BigTechs - will be able to benefit from accessing the EU market and will not be held accountable for the detriment that AI technologies might cause to EU residents. This situation represents a major imbalance of market power between the technology providers and customers. It bears uncalculated risks for society, as the unpredictability of future AI developments will be caused, among others, by the lack of accountability regime. **The EU Commission should table a new proposal to establish non-contractual civil liability rules for AI.**

Oddly, Article 14.4.b. of the EU AI Act covers the case of the duty of advice and the risk of ill-founded financial advice for high-risk systems defined when it comes to financial services in Annex III.5 (c) as *"AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance"*. There is no doubt that life and health insurance should be in the scope of the EU

AI Act, and this is a good point, but limiting this scope to life and health insurance products creates an obvious loophole in the regulation, as it makes human oversight compulsory for those products but not for financial products in general. This makes regulation and, subsequently, supervision of financial products dependent on their legal wrapping, and thereby creates regulatory arbitrage possibilities: with the current wording of the EU AI Act, the same financial product with the same risk profile sold to the same customers (natural persons) will, for instance, require human oversight if packaged as a life insurance contract but not if packaged as a UCITS or a structured deposit.

Challenge n°3: AI's inherent lack of transparency makes the output of AI models largely impossible to explain.

AI models' efficiency is rooted in their ability to process enormous amounts of structured and unstructured data which, by definition, makes them non-transparent and makes their results difficult, if not impossible, to explain. This in-built opacity leads to a "take my word for it" logic that runs against the principles of transparency and explainability.

Human oversight can detect extreme cases, the oft-discussed hallucinations, but it is unable to detect the vast majority of plausible but false claims that an AI model can make. AI models routinely make such plausible but false claims due to the way they operate, including their inability to say "I don't know". Large language models do not function on an "I know" / "I don't know" logic (this is the human way); they are trained to infer plausible answers. In most cases, the plausible answers provided by LLMs happen to be correct. However, the problem arises when LLMs come up with plausible but false assertions. In the majority of cases, the human receiving the AI-produced statement will not be in a position to detect its falsehood, and in the absence of transparency the situation is unmanageable and potentially deleterious.

This is a fundamental difference with the quantitative approach that has been used for many years in financial activities. However complex and abstruse they may be for the non-specialists, the now traditional models used in finance are transparent and explainable for their users. For instance, option pricing models, whether analytical (e.g. Black Scholes and its many variations…) or numerical (e.g. Monte Carlo, finite differences…), two very different mathematical approaches to solving the question of the fair value of an option, can be trusted by derivatives professionals as they understand how they run, and therefore what their assumptions and their limits are. By the same token, banks' asset liability management teams can and do share with supervisors the sophisticated internal models they use to calculate risk weighted assets, and supervisors understand how they work, at least in principle. The confidence that professional users have in those models derives from the fact that they know their limits and, therefore, when – and when not – to trust them. Said differently, the usability and the usefulness of even complex mathematical models in finance comes from their characteristic of transparency for the professionals using them. Those models enable professionals to make explainable decisions, even if often based on an ill-founded sense of security linked to the use of sophisticated mathematical techniques.

This is in sharp contrast to AI's approach which feeds a vastly blind decision-making process in which human oversight is only able to filter out extreme cases (so-called **hallucinations**). This "trust me blindly" characteristic of AI models is based on the claim that they have seen every single possible state of the world during their training. Confronted with the output of an LLM, users find themselves in a take-it-or-leave-it position and they end-up taking it, unable to challenge it. This phenomenon was described recently by an authoritative voice on the subject when Dario Amodei, CEO of Anthropic, one of the leading AI firms and developer of Claude LLM, said in an interview[10] that "*maybe now we understand like 3% of how they (AI models) work*" and "*we don't understand how all of these things interact to give the behaviour that we see from models everyday*". To say the least, this dynamic is not the best way to create trust, which creates an obvious problem in the world of finance – where **trust** is of the essence.

This lack of trust and transparency has direct implications in the financial sector, as illustrated by the following examples:

1. In the field of anti-money laundering already mentioned, how can a financial institution compliance officer challenge the result of its AI tool indicating no suspicious transfer of money?
2. In the field of financial research, how can an analyst challenge a wrong, albeit plausible, inference about a corporation?

The latter question (wrong inference) is particularly important given that LLMs, by construction, do not know how to say "I don't know". How can one act on the basis of the output provided by an AI model if false responses are undetectable by a human eye?

AI does not have a black box problem (implicitly "that could be solved"). It is by construction a **black box**, and moreover a black box whose results even its creators, by their own admission, do not understand. This is most problematic when it comes to using AI in the field of financial services or making financial decisions, in particular when the percentage of AI models' outputs we do not understand reaches 97%, as stated by Anthropic CEO Dario Amodei.

Challenge n°4: towards an exhaustion of human-generated data leading to a collapse of AI models?

EPOCH AI summarised a ground-breaking paper[11] published in June 2024 with the following introductory remark: "*We estimate the stock of human-generated public text at around 300 trillion tokens. If trends continue, language models will fully utilize this stock between 2026 and 2032, or even earlier if intensely overtrained*".

---

10 *In Good Company*, Nicolai Tangen, from 2:43-3:31.

11 Villalobos, P., Ho, A., Sevilla, J., et al. '*Will we run out of data? Limits of LLM scaling based on human-generated data*'. ArXiv [cs.LG], 2024. arXiv.

A month later (July 2024), the authors of an equally important paper[12] described how "*AI models collapse when trained on recursively generated data*", summarising the finding of their research in the following way: "*We find that indiscriminate use of model-generated content in training causes irreversible defects in the resulting models, in which tails of the original content distribution disappear. We refer to this effect as 'model collapse' and show that it can occur in LLMs as well as in variational autoencoders (VAEs) and Gaussian mixture models (GMMs). We build theoretical intuition behind the phenomenon and portray its ubiquity among all learned generative models*". In other words, **AI models collapse** (i.e. start providing nonsensical results) when trained on non-human-generated / synthetic data, a phenomenon referred to by specialists as loss of variance.

Linking the two aforementioned papers, we can see appear what may be one of the biggest challenges of AI: an exhaustion of human-generated data leading to a collapse of AI models, a situation described as a "bottleneck" by Anthropic CEO Dario Amodei in the interview cited above.

This phenomenon only reinforces challenge n° 3 (lack of transparency leading to lack of trust) and makes the situation even more problematic when it comes to financial services. Not only are humans often unable to detect the veracity, or lack thereof, of AI-produced outputs, but we face a situation where human-generated data will be exhausted within a limited number of years. When that happens AI models are likely to produce false or meaningless results, errors that may not be detectable by a human eye. This poses yet another challenge for developing the trust necessary for the functioning of an increasingly AI-powered financial system.

Challenge n°5: the risk of data manipulation.

Data manipulation is another data-related challenge arising along the non-human-generated / synthetic data problem, leading to nonsensical results (challenge n°4) and the un-detectability of falsehood (challenge n°3).

AI models developed or used specifically for the provision of financial services are, or will be, trained on specific datasets. This is in the nature of AI. Evidently, if finance-specific AI models are trained on **biased or manipulated datasets**, they will produce **biased or manipulated results.** The manipulation of the training data set can take different forms: incompleteness, selection biases, the choice of a deliberate "angle" suiting a specific interest, and/or the intentional injection of data. Models are only models and biased data sets will result in biased model outputs. For instance, the risk in the field of financial services to retail customers is, for instance, to see the AI models trained on data with a "maximise the profits / customer interest comes second" bias. Biased datasets can be the result of malicious action or traditional business interests but, regardless of the cause of the bias, the consequence will be the same: biased

12 Shumailov, I., Shumaylov, Z., Zhao, Y. et al. *AI models collapse when trained on recursively generated data*. Nature 631, 755–759 (2024) (op. cité).

outputs and unsatisfactory, if not detrimental, decisions for the users of AI models' results.

In its AI governance paper published in 2021,[13] EIOPA notes on page 14 that "*In a fast-moving digital world, there is a strong imbalance between those who manage algorithms and data, and the data subjects, the latter struggling to exercise their rights*". This issue is not unrelated to the asymmetry of information theories developed some 50 years ago in a non-AI age by Akerlof, Rothschild and Stiglitz.

B. Reconciling AI with the EU financial regulation rule book

a. Context

Most papers dealing with the use of AI in finance give a list of challenges to address. For instance, the Bank for International Settlements (BIS)[14] gives the following list: "*Black box mechanisms, algorithmic discrimination, zero-sum arms races, model herding, algorithmic coordination, algorithmic collusion, new liquidity crises, increased cyber risks, hallucinations, increased market concentration*". To this list can be added consumer privacy concerns, customer profiling, product design, inappropriate financial advice, as well as the lack of accountability, responsibility, and transparency. Finally, there is the fundamental question of "how can a financial system, relying in essence on trust, function on the back of an engine (AI) that even its most advanced developers do not fully understand?"

AI does not only introduce challenges and risks. To a significant extent, **AI contradicts** some of the most **fundamental principles** underlying the financial system and existing financial regulation.

This is particularly obvious when examining the **question of accountability**. As described above in "Challenge n°2: is the provision of AI-powered financial services above the law?", the entire financial regulation framework is based on the principle of accountability, which is itself based on the possibility to attribute activities to a natural or legal person to establish liability. However, this principle is, in many cases, not compatible with the principles of AI and its combination of processing power beyond the capability of the human brain and inductive process.

When the BIS states[15] *"Thus, effective regulation and governance frameworks are important to harness the benefits of AI while mitigating associated risks, emphasising transparency, fairness and global collaboration"*, it sets in our view the right objective – effective regulation governance frameworks – but lacks realism suggesting mitigating associated risks, emphasising transparency, fairness and global collaboration on the agenda. By definition AI is not transparent; it is by construction a black box. Fairness

13 *Artificial intelligence governance principles*, EIOPA, June 2021.

14 BIS Working Papers, No 1194 – *Intelligent financial system: how AI is transforming finance*, June 2024.

15 Ibid, page 31.

is not its concern and collaboration even if obviously desirable does not seem to be many people's priority these days.

Traditionally, finance has been, and to a large extent still is, about providing financial services in accordance with the rules set by the law and enforced by supervisors. It has gone mostly unnoticed that this is a fundamental difference between financial regulation and the regulation of other industries. To illustrate this difference, the process used by the chemical industry to produce a certain product is not changed by a regulation preventing it from disposing of hazardous materials into local rivers. To the contrary, financial regulation actually shapes the way the financial industry operates, the type of financial products it designs and sells, etc. Financial regulation, contrary to other industries' regulation, is performative: the rules are the modus operandi of financial activity and they influence the very nature of the services on offer.

However, AI changes this landscape in quite a dramatic manner. When governed by AI, financial services are no longer a function of setting existing financial regulation to music, but of doing it the AI way. Attracted by the potentially enormous efficiency gains and subsequent increased profits, the financial industry effectively frees itself from this traditional logic, betting on the fact that supervisors will not dare intervene on the back of the mantra of the impossibility to stop innovation. **Fait accompli policy**.

Those issues are central to the endeavour of making AI and financial regulation compatible, which is de facto an ex-post approach (regulation is struggling to catch-up with the practices encountered in the financial sector).

b. A concrete method to reconcile AI and financial regulation

In the EU, as in most jurisdictions, the provision of financial services is regulated: an authorisation from the relevant National Competent Authority (NCA) is necessary to sell financial services and it is granted on the basis of a strict compliance of the newly authorised activity with the regulation rule book.

In principle, all jurisdictions apply a "**same activity, same risk, same rule**" principle, which on paper insulates the financial regulation rule book from the risk of having to evolve each time a service provider innovates on the manner it provides its services to its customers or conducts its business. If the objective or the end result are the same, then the same rule must apply.

Two questions arise when it comes to questioning the possibility of regulating the provision of AI-powered financial services:

1. Can the services provided be considered as being the same?
2. Is the regulation of those services technically possible?

The first question (knowing whether the services provided can be considered as being the same) must be assessed from a teleological standpoint. In most cases, if not all, the response will be yes. Judging from the objective (result), activities as different as

providing financial advice, profiling customers, optimising capital allocation, assessing credit quality or detecting suspicious transfers of money can be considered as being the same, regardless of whether the activities are managed by human or artificial intelligence. By and large, this is the case for most financial services and activities.

In contrast, the answer to the second question (knowing whether the regulation of those services is technically possible) is not as clear-cut: in some cases, yes, and in some cases, no. Regulating an activity is an exercise of attributing the responsibility of an outcome to an individual or to a legal person. In the cases where such a responsibility can technically be attributed (e.g. financial advice) then AI-powered financial services can and must be regulated in the traditional way.

However, in the cases where no such responsibility can be attributed because of the nature of the service and as a consequence of the "**correlation is not causation**" principle, the very idea of regulating contains a contradiction when decisions are made on the sole basis of AI. In those cases, policy-makers are faced with a three-pronged possibility: 1) prohibit the AI-powered activity or service; 2) let it develop unregulated; 3) regulate it in a different way.

With the conviction that, like it or not, prohibiting is not realistic and that letting a financial activity develop unregulated is undesirable, Finance Watch proposes a novel way of regulating and supervising those activities, admittedly a second best compared to the traditional and excellent responsibility principle, but still better than nothing.

We describe this second best method as regulating and supervising AI-powered financial services at data and **data governance level**. In other words, when responsibility of the outcome cannot be attributed, regulating and supervising the beginning of the chain, i.e. data and data governance, is a better option than either developing an illusory regulation of the outcome that will not be applied in practice, and obviously better than not regulating at all.

However, this principle should be handled with great care: not only should it be applied only to the cases where responsibility is not technically attributable because of the specificities of AI, but in the cases where allowing the provision of AI-powered financial services without the possibility of attributing the responsibility of the outcome presents a threat for essential public interests, such a provision should be prohibited altogether.

For instance, in the concrete examples already taken in this report, we would argue that providing financial advice to private individuals, or controlling the legality and the compatibility of financial flows with the Anti-Money Laundering Directive, must always be dependent on the possibility of attributing responsibility of the outcome to a legal or a natural person (or to both). Establishing non-contractual **civil liability rules** for AI is

a necessary step in this direction. However, notwithstanding such liability provisions,[16] it is necessary to assess the possibility for supervisors to enforce existing rules for AI-powered financial services. Alternatively, regulating and therefore controlling only at data and data governance level, though a second best, could be considered for use cases linked to risk-management. Distinguishing between those two blocks should be the first focal point of the gap analysis to be conducted. This will be essential to support the work of legislators who will have a delicate balance to find when reviewing existing financial regulation to ensure it is still fit for purpose in an AI world.

16 As mentioned in the preceding chapters, the European Commission has withdrawn its proposal for the AI Liability Directive (AILD).

III. Adapting EU legislation to the provision of AI-powered financial services

A. The EU AI Act as the start of the journey

The EU AI Act,[17] adopted in June 2024 and becoming applicable in August 2026, has an objective of addressing the challenges of AI. In order to do so, it establishes a risk-based approach with four levels of risk identified for AI activities and different rules applying to each level:

1. Unacceptable risk
2. High risk
3. Limited risk
4. Minimal risk

The AI Act is articulated to a large extent around the classification of high-risk AI systems given in Article 6 and the requirements applying to those systems (Articles 8 to 15).

Financial activities are only partially covered by the EU AI Act's Article 6 (implicitly) and in Annex III (explicitly). We argue that if the AI Act's approach to financial services has merits, its scope needs to be substantially broadened for the Act to become meaningful for the financial sector. Such a broadening of scope would allow for the important provisions of Articles 8 to 15 to apply to financial activities, among which Article 9 (Risk Management System), Article 10 (Data and Governance) and 14 (Human Oversight) are of particular importance for the field of finance.

Article 6 and Annex III of the AI Act, or parts of them, deal with financial services.

Financial services in the EU AI Act: relevant extracts of Article 6 and Annex III

Article 6: Classification Rules for High-Risk AI Systems

*3.(d) (…)Notwithstanding the first subparagraph, an AI system referred to in **Annex III** shall always be considered to be high-risk where the AI system performs profiling of natural persons.*

Annex III: High-Risk AI Systems Referred to in Article 6(2)

High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

17 *Text of EU AI Act*. *Context of EU AI Act*.

(…)

5. Access to and enjoyment of essential private services and essential public services and benefits:

(…)

(b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud;

(c) AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance;

Considering as high-risk the evaluation of the creditworthiness of natural persons is the right approach, for the reason that such an evaluation is by nature a profiling exercise, which raises important **privacy and discrimination issues**. By the same token, risk assessment and pricing in relation to natural persons in the case of life and health insurance is high-risk, as it leads to the question of the appropriateness of the financial advice provided to natural persons and bears financial exclusion risks based on discrimination. However, those provisions need to be applied to a much broader range of financial services for the very same reasons: there is no justification to subject an investment product in life insurance wrapping to the AI Act when the same product in different wrapping is not, and such a provision will mechanically trigger **regulatory arbitrage** to the detriment of life insurance and retail customers. Similarly, risk assessment and pricing is not only high-risk when done in relation to natural persons: when done in relation to financial institutions it is also high-risk, if in a different manner, as it concerns financial stability which is a public good.

AI has implications on the way all financial activities are conducted. This should lead us to include all financial activities in Annex III of the EU AI Act, define them as **high-risk** and subsequently see the provisions of Articles 8 to 15 apply to them. However, we have established that human oversight will not always be technically possible regardless of its desirability and, as a consequence, a carve-out focused on Article 14 (Human Oversight) could be considered for the services where, after a thorough review, not attributing responsibility can be seen as acceptable. The corollary of such a carve-out should be the prohibition of those AI-powered financial services for which human oversight is impossible but responsibility is indispensable. When extended to cover all financial services, the AI Act should list the financial services for which human oversight is possible, and those for which it defies the very purpose of going the AI route. In the latter cases, data and data governance (Article 10) would become the most important and effective line of defence of the public interest on the rationale that unbiased and quality data constitute the prerequisite for producing quality outcomes

in an AI world. This is true regardless of the specific dimension taken by the public interest for each financial activity contemplated, and it makes a strong case for applying Article 10 of the AI Act to all financial services.

In terms of process, we propose to broaden the scope of financial activities covered by the EU AI Act by activating the provisions of Article 7 (Amendments to Annex III) which gives the European Commission the power of amending Annex III through the adoption of delegated acts. The best approach would consist of broadening the scope of ANNEX III, 5. (b) and (c) to cover all financial activities, which would mechanically subject them to Articles 8 to 15 of the AI Act, possibly with the caveats or exemptions already mentioned for AI-powered financial activities not compatible with human oversight. Article 5 of the EU AI Act (Prohibited AI Practices) should also be amended to include financial services for which human oversight is not possible and not attributing responsibility is not acceptable.

Article 7 of the EU AI Act

Article 7: Amendments to Annex III

1. The Commission is empowered to adopt delegated acts in accordance with ***Article 97*** to amend ***Annex III*** by adding or modifying use-cases of high-risk AI systems where both of the following conditions are fulfilled:

2. When assessing the condition under paragraph 1, point (b), the Commission shall take into account the following criteria:

(a) the intended purpose of the AI system;

(b) the extent to which an AI system has been used or is likely to be used;

(c) the nature and amount of the data processed and used by the AI system, in particular whether special categories of personal data are processed;

(h) the extent to which there is an imbalance of power, or the persons who are potentially harmed or suffer an adverse impact are in a vulnerable position in relation to the deployer of an AI system, in particular due to status, authority, knowledge, economic or social circumstances, or age.

B. Beyond the EU AI Act: the case for reopening existing financial regulations

Broadening the scope of the EU AI Act to cover financial services in a comprehensive and effective manner, and adopting a non-contractual civil liability regime for AI setting in stone a '**presumption of causality**' are a necessity. However, even with a broadened scope, the AI Act as it has been designed will not be sufficient to protect the interest of financial services users, investors and citizens.

With this objective in mind, a gap analysis will have to be conducted to determine which regulations must be reopened and what amendments must be

made to ensure that the interests of investors, consumers, citizens and society at large are protected in a world of AI-powered financial services.

Conducting a detailed gap analysis is beyond the scope of this report. As a first step, the work done here provides a list of the themes and legislations that must be covered. It also puts them in perspective with the specific vulnerabilities and challenges created by the intrusion of AI in the provision of financial services and the conduct of financial activities.

1. Themes
   - → Suitability and appropriateness of financial services and financial products
   - → Fairness and non-discrimination
   - → Privacy, data governance and data protection
   - → Transparency and explainability
   - → Human oversight
   - → Pricing of financial products / price optimization practices
   - → Accountability and legal redress
   - → Market manipulation
   - → Financial stability
   - → Security and systems

2. Legislations
   - → A non-exhaustive view of the legislations relevant for a gap analysis:

	Suitability / appro-priatness	Fairness	Privacy and data	Transparency	Human oversight	Pricing	Accounta-bility	Market mani-pulation	Financial stability	Security and systems
EU Primary law										
Treaties		X								
Charter of Fundamental Rights		X					X (Art. 47)			
EU secondary law										
Insurance Distribution Directive	X (Art. 30)			X (Art. 20)	X	X				
MiFID II	X (Art. 25)			X	X			X (Art. 17)	X	
Capital Requirements Directive (CRD)				X	X	X			X	
Solvency II				X	X (Art. 41)	X			X	
General Data Protection Regulation (GDPR)			X	X (Art. 5, 6, 13 and 14)	(Art. 37)					
Anti-discrimnation Directives		X	X							
Unfair Commercial Practices Directive (UCPD)		X				X				
Consumer Credit Directive (CCD)		X (Art. 18)	X (Art. 18, 19)	X (Art. 18)	X (Art. 18)					
Distance Marketing of Financial Services (DMFS)	X	X	X							
Digital Operational Resilience Act (DORA)			X						X	X (Art.4)
Anti-Money Laundering Directive (AMLD)		X (Art. 13-18)		X (Art. 13-18)	X (Art. 13-18)					
Investment Firms Directive (IFD)	X (Art. 36)									
Market Abuse Regulation (MAR)				X				X (Art. 12, 15)		
Mortgage Credit Directive (MCD)	X (Art. 22)	X (Art. 18)	X (Art. 18)	X (Art. 18)	X (Art. 18)					
UN Treaties										
Human Rights treaties		X	X							
Council of Europe										
European Convention on Human Rights		X	X							

3. Vulnerabilities

The analysis of the vulnerabilities that arise in the field of financial services when they are provided using artificial intelligence should be the guiding thread of the detailed **gap analysis** to be conducted and of the subsequent possible reopening of existing financial regulations.

With that logic in mind, this analysis examines the vulnerabilities created by AI in the fields of consumer protection, risk management and financial stability, and lastly compliance, supervision and fraud detection.

- In the field of consumer protection:
  - → Fundamental vulnerability:
    - ✓ **Consumer protection** is the protection of every single customer (with sanctions in case of faulty compliance), not the protection of consumers on average. This means that even in situations where all the intentions of the financial services provider are legitimate, and the data used to train the AI models is as unbiased as can be, the provision of AI-powered retail financial services can, and will, "get it wrong" in specific cases. This state of affairs has broad implications for financial institutions who will have to choose between the increased efficiency (and therefore profitability) brought by AI and the cost of remediation to provide to the customers whose interest will not have been taken care of. In any case, financial regulation has to ensure that financial institutions remain legally responsible in the field of consumer protection.
  - → Other specific and often-encountered vulnerabilities or challenges:
    - ✓ AI-linked **price optimisation** practices (e.g. "willingness to pay" and other techniques penalising loyal customers).
    - ✓ Developing **non-discriminatory** machine learning and in particular "fair machine learning" (i.e. ML going beyond removing biases from the training data).[18]
    - ✓ **Transparency and explainability** for customers of AI-powered retail financial services.
    - ✓ Making financial services providers **accountable** in case of mis-selling, non-suitability or non-appropriateness of the investment products sold to retail customers.
    - ✓ Ensuring AI-powered pricing of insurance products operating away from the traditional pooling / mutualisation logic does not jeopardise the societal role of insurance, and making it non-discriminatory.

18 It should be noted that the possibility of "fair machine learning" is a subject of non-consensual debate among specialists.

- In the field of risk management and financial stability:
  - → Fundamental vulnerability:
    - ✓ Among the risks of AI when used for risk management, investment or trading purposes, **herding**, stemming from generalised use of the same model, is often pointed out as a major risk. The logic is that when all actors behave in the same manner as their models and send the same buy or sell signals at the same moment, pro-cyclicality kicks in, which can lead to market stampedes. This risk, sometimes dubbed correlation risk, is undoubtedly real but it is not related to the existence of AI.
    - ✓ Herding is hardly anything new in the history of financial crises. It is also the essence of speculation. If fear and greed (both the result of herd instinct) are the eternal engines of financial markets' momentum, herding based on the widespread use of mathematical models has existed for at least 40 years, witness the market crash of 1987 in a pre-AI world. Herding is, however, only made a bigger source of instability by AI for two reasons. Firstly, the famous beauty contest analogy used by Keynes to describe speculation in chapter 12 of its "General Theory" is compounded by the use of AI to second guess, third guess, etc. the average market opinion itself produced by other algorithms. Secondly, the financial system faces a growing risk of "**squared oligopoly**", a concept coined in the book "Sopravvivere nell'era dell'intelligenza artificiale"[19] to describe a situation combining a limited number of financial giants managing ever larger amounts of financial assets and the limited number of companies providing the AI algorithms. The squared oligopoly situation increases the risk of market stampedes and subsequent financial instability.

      Beyond the question of herding, the main question is the **overconfidence** human beings tend to have in the results produced by models and that the risk coming with this overconfidence will only be reinforced as the usage of AI models for risk-management purposes becomes more and more widespread. The story often heard on the subject is that, with its enormous computing power, AI is capable of detecting previously undetectable patterns and therefore of predicting the future.[20] This assertion is deeply flawed as it infers that the condition necessary to predict the future could be to process in a more effective way existing (by nature backward looking) data, thereby making the assumption that the future will be a repetition of the past.

      Financial risk management can be described as the hunt of black swans, i.e. of events outside of the traditional probability distributions (normal,

19 "Sopravvivere nell'era dell'intelligenza artificiale", collective book, 2024.

20 See, for instance, BIS Working Papers No 1194, Intelligent financial systems: how AI is transforming finance - page 10 - "*ML is also heavily used in asset pricing, in particular to predict returns, to evaluate risk return trade-offs, and for optimal portfolio allocation*".

log-normal, Poisson…). Said differently, chasing statistically impossible events that happen nonetheless in the real world.

We argue that LLMs' and GenAI models are not better equipped than traditional probabilistic models to detect **black swans** and avoid them, given that their outputs are the result of a stochastic process. Being an exercise of minimisation of the average statistical error, AI recognises implicitly that errors exist (that black swans exist). By definition, inductive reasoning based on past data cannot detect or predict future extraordinary events, i.e. events not encountered in past patterns. The fact that AI-powered risk-management applications are the 'product' of the dataset they were trained on raises the two classical questions Bertrand Russel raised about induction: 1) Is the sample used wide enough? 2) Will the future look like the past? Arguably, AI might be considered as resolving the first question but in no case can it resolve the second one.

Using AI models as risk management tools comes with vulnerabilities as it further reinforces the ill-founded sense of security given to (bad) risk-managers by models. This ill-founded sense of security gets even more dangerous in the current context of an unsustainable economic activity and a world fraught with radical uncertainty. AI models are no more equipped to detect **green swans** than they are to detect black swans. Their ability to predict future returns and crises is equal to their ability to write a poem "à la manière de" of a yet unborn poet who, by definition, has not yet written anything.

AI is useful and powerful as long as the future looks like the past. For many fields, for instance scientific research, it makes it an incredibly valuable tool, if not without its idiosyncratic challenges. However, for financial risk management purposes in a radically uncertain and fast evolving world, it makes it a dangerous tool if blindly relied upon. Human risk managers using AI tools need not forget that the expected value of an event (i.e. its probability multiplied by its impact) comes with strong limitations when used for risk management purposes. This is due not only to the fact that 25 standard deviation events can happen several times a day during financial crises but, even more importantly, because a responsible risk manager cannot take a 1 in 10,000 risk if the occurrence of that risk is synonymous with the bankruptcy of the institution he or she is in charge of.

→ Other specific and often-encountered vulnerabilities or challenges[21]:

✓ **Third-party dependencies:** for both cost and efficiency reasons, AI professional users, including financial institutions, tend to rely on third-party service providers offering hardware and cloud services, pre-trained mo-

21 This section takes inspiration, even though in a non-exhaustive manner, from the FSB 2024 report on "The Financial Stability Implications of Artificial Intelligence".

dels and financial data aggregation. Given the relatively limited number of such providers and their largely unregulated nature, this creates a major vulnerability for the financial sector.

✓ **Cyber risk:** the generalized adoption of AI is a significant factor of enhanced cyber risk at all levels of the financial sector. As described by the Financial Stability Board,[22] "*LLMs and GenAI could enhance cyber threat actors' capabilities and increase the frequency and impact of cyber attacks. Intense data usage and novel modes of interacting with AI services increase the number of cyber attack opportunities. Greater usage of specialised service providers exposes FIs to operational risk from cyber events affecting these vendors*".

✓ **Model risk and data quality:** before the intrusion of AI, the models used in finance were either developed by the financial institutions' specialised staff (the so-called "quants") or purchased from vendors providing transparency (no black box). They were understood by the risk departments of the institutions and they were tested internally before they were put in operation. With LLMs and GenAI, the situation is entirely different: we now evolve in a world where, as described by Anthropic CEO Dario Amodei (cf. supra), "*we understand 3% of how AI models work*". This is not the same game and it spells trouble when it comes to using those models for risk management purposes. The data issue has already been described (Challenge n° 4: towards an exhaustion of human-generated data leading to a collapse of AI models?). It is not only a decisive factor for the quality and the adequacy of AI-powered risk management models but also an unresolved problem in a context of foreseen primary (human generated) data exhaustion and the subsequent possibility of AI models collapse (cf. supra).

- In the field of compliance, supervision and fraud detection:

  → Fundamental vulnerability:

  ✓ The intrusion of AI in the world of rule-abiding is the story of an arms **race between regulated entities, supervisors and fraudsters**. AI makes the three categories of actors more efficient and powerful. AI has now made its way into the compliance processes of regulated entities (Regulatory Technology / RegTech), thereby improving their ability to monitor the compliance of their operations with a large and often complex web of financial regulations. It has also entered the world of supervisors (Supervisory Technology / SupTech), gradually enhancing their supervisory capability by an order of magnitude.[23] However, it is also becoming the favourite tool of financial fraudsters, to a point where

22 Op. cit.

23 According to the Cambridge Centre for Alternative Finance (2023), Cambridge SupTech Lab: *State of SupTech Report 2023*, 59% of financial supervisory authorities were using SupTech in 2023.

the Financial Stability Board stated in November 2024[24] that "*AI can help FIs and authorities fight fraud, but GenAI may benefit malicious actors more than legitimate actors in the short run". "Thus, in the near term, it will likely be easier to generate fraudulent content using GenAI than to detect it*".

→ Other specific and often-encountered vulnerabilities or challenges for supervisors:

- ✓ Detecting **deep fakes** and preventing disinformation leading to manipulations potentially triggering market crashes, bank runs or liquidity crises.
- ✓ Developing a framework to vet / authorize only models that can be explained in complete **transparency**. This will be a particular challenge for AI models used to compute assets' risk weights and therefore banks' capital requirements for Basel III / Capital Requirements Regulation internal ratings-based purposes.
- ✓ **Supervising the learning process**, including the quality of the data used for training the models in a context where the quality of AI models' output is dependent on the quality of the data used to train the models.
- ✓ Ensuring they can enforce existing rules for AI-powered financial services.
- ✓ Finally, as outlined by the FSB p. 29 of its 2024 AI report, "*financial authorities face two key challenges for effective vulnerabilities surveillance: the speed of AI change and the lack of data on AI usage in the financial sector*".

Ethics and fairness in insurance[25]

The concept of fairness plays an important role in classic liberal thought as a manner of providing common ground for the resolution of conflict. In other words, fairness can furnish societal actors with a neutral space for the management of disputes. The idea of fairness is closely related to other key concepts that underpin any society such as justice, equal opportunity, freedom, trust, responsibility and accountability.

24 FSB, The Financial Stability Implications of Artificial Intelligence – 14 November 2024, page 26 (op. cite).

25 Box page 12 and 13 of EIOPA's report '*Artificial intelligence governance principles: towards ethical and trustworthy artificial intelligence in the European insurance sector*' of June 2021.

The idea of fairness spans multiple disciplines including sociology, law and politics and relates to the goal of the equal treatment of citizens and non-discrimination. That said, how fairness operates or *what is fair* remains highly contested and with much of the debate centring on the relationship between fairness and equality. Moreover, fairness has been linked to the concept of deservedness and the notion that people get what they deserve according to such attributes as hard work or particular skill-sets.

Within John Rawls' seminal work "A Theory of Justice", fairness plays a key role in the establishment of the "original position" in which citizens decide on the future shape of society without knowledge of their position in that society. Although not related to public policy-making in its widest sense, we can detect some resonance with the paradigm of risk-sharing and the practice of insurance in that on becoming part of the risk pool, participants are unaware of whether or not they will require compensation for an adverse event. Like insurance, fairness has an important temporal element as fair outcomes and risk profiles for citizens relate to life course and change over time. This is an important and often overlooked element as insurance provides security through different ages and is there to deal with changing fortunes.

In insurance, notions of fairness need to capture the interests of insurance firms, individual insured customers, the pool of insureds, and society as a whole. Their interests will impact how the concept of fairness is defined so, for example, insurance firms may stress their right to conduct the business of insurance freely within the legal bounds. Representatives of individual insureds may define fairness in this market as *inclusivity*. Representatives of the pool of insureds may stress actuarial fairness, according to which similar risks are treated similarly, so that the premium paid by individuals corresponds to their actual risk, taking into account that there are other factors that influence the premium (e.g. production costs). Society as a whole may put an efficient, well-functioning insurance market at the centre of its interests, as this fosters welfare and economic activity. The subject of fairness, responsibility and digital ethics in insurance markets has attracted a good deal of attention within the academy. A number of recently published papers attempt to deal with the issues around fairness and the use of AI by insurers.

Many of the elements of the broad concept of fairness are reflected in existing professional practice and insurance and data protection regulation. The term "fairness" relates to requirements concerning the business conduct of insurance firms towards consumers. This includes policies on non-discrimination, access to insurance and the treatment of vulnerable consumers. Paradoxically, digitalisation represents both a challenge to establishing fairness in insurance and provides a means to implement more fairness in insurance in the future.

In the pursuit of fairness, some reflection is needed on whether it is “natural and inevitable” that the interests of stakeholders are inevitably opposed. If we consider the insurance firms and policyholders - the interests of both are intrinsically linked, the former cannot exist without the latter and the latter cannot have peace of mind without the former. The fact that the consumer seeks the best value for money and insurance firms seek to make sure the economic expectations of their owners are met (whether shareholders, investors, customers (e.g. in mutual insurers), or other stakeholders) means that their interests may not be aligned, but they also cannot be described as completely opposing. The concept of fairness also pertains to competition between insurance companies.

Fairness in insurance is related to how the insurance market operates for society. Hence the fair operation of an insurance market is one where insurance firms offer reliable and effective insurance that is easy to compare and is marketed in a way that consumers can make informed choices. It would also be a market that provides insurance products that are essential for society. That could include a range of simple and affordable, default option premiums to afford unlucky and/or vulnerable people reasonable access. The notion of “essential insurance” is central here and speaks to challenges and demands from citizens/customers and public authorities. Here we see the important role that insurance plays in terms of societal responsibility and social equity in co-creating a market that offers opportunity and access to all citizens including vulnerable groups, to access essential products at an affordable price. On the other hand, unfortunately it is not always possible for insurance firms to offer insurance at an “affordable price” (e.g. flood insurance in heavily exposed areas, terrorist risk, pandemic risk etc.). In such cases it is a political and societal task to find solutions

Conclusion

Many questions arise in the wake of the development of artificial intelligence. Among those, three stand out as particularly significant : 1) the enormous energy consumption necessary to train and run AI models at a time when decreasing power consumption should be the highest priority of mankind for its very survival, 2) the possibility of mass manipulation coming with the generalisation of deep fakes, and 3) AI's monopoly problem and self-reinforcing concentration.[26] In an ideal world, these three questions would have been debated before the development of AI.

Regulating AI is an exercise of finding the right balance between the irrepressible business interests behind the AI momentum and the public interest. AI is in fact a showcase, if ever there was one, that public interest is not the sum of private interests. The case of private interests having been made over and again for AI, forgetting public interest on the back of private sector profitability or economic competition arguments can only lead to societal disasters. This, in turn, will be detrimental not only to citizens and to society, but also to business interests themselves.

Enforcing the rules is as important as adopting the right rules, and supervising AI-powered financial services in an effective manner will be of vital importance. This will imply giving the means to supervisors to operate with efficacy, in other words giving them the financial means of recruiting AI specialists and investing in adequate computer power. This might prove a challenge in times of constrained public finances, which makes a strong case for raising the necessary revenues from the (huge) winners of the AI boom, in other words from AI services providers themselves. The alternative – namely not equipping supervisors properly to control the provision of AI-powered financial services – would spell disaster for financial services users, for financial services providers and for society.

This report has made the case for regulating the provision of AI-powered financial services. Undoubtedly, the anti-regulation camp will raise the usual 'stifling innovation' and other lobby-manufactured arguments along the lines of 'the EU is only good at regulating while the US innovates'. EU policy-makers should not be impressed by such empty and predictable rhetoric unless they think that innovation is an objective in itself (we don't: we see innovation as a means to an end, not as the end) and that public interest can be equated to the wealth of a small number of players without regard for the interest of society, citizens and consumers. At the heart of AI development is also a battle between a handful of libertarian multibillionaires with a dream of becoming ever more powerful, and the society of normal citizens representing 99.9% of mankind.

AI is here to stay. It is a most powerful tool and it brings with it benefits in a number of fields (e.g. scientific research, white collars' work…) … and great societal evils

26 See for instance: Open Markets Institute – AI in the Public Interest: *Confronting the Monopoly Threat*, November 2023.

(e.g. deep fakes, mass manipulations, the possibility of widespread discrimination, increased fraud, an absurd level of energy consumption at a time when saving energy should be mankind's highest priority…). Things being what they are, the objective is now to take AI's benefits and minimise its evils. This will undoubtedly be a great challenge given the forces at play.

In the field of finance where AI is entirely about increased productivity and profitability of service providers and financial institutions, policy-makers must introduce rules to protect the public interest. A certain conception of society is at stake where public interest remains the mission of policy-makers and private interests, however natural and legitimate, are not necessarily king by their sole money-making virtue. Letting greed and the will to power of a few go unfettered is not a good policy option.

ANNEX I

Responses provided by OpenAI's ChatGPT and Google's Gemini to the question: "*What are the EU financial regulations impacted or potentially impacted by the use of AI in financial services?*"

The question was asked to ChatGPT and Gemini on 28 November 2024 after this report was finished writing. It therefore did not have any influence on its content.

The responses to the question have been copied and pasted without any editing, thus enabling readers to judge by themselves the quality of the responses provided by the two LLMs.

ChatGPT's response:

The use of artificial intelligence (AI) in financial services intersects with a variety of existing and emerging European Union (EU) financial regulations. Key areas of impact include governance, consumer protection, risk management, transparency, and ethical considerations. Below are the main EU financial regulations that are impacted or potentially impacted by AI in financial services:

1. **General Data Protection Regulation (GDPR)**
   - → **Relevance to AI:** AI systems in financial services often rely on large datasets, including personal data, for training and decision-making.
   - → **Key Impacts:**
     - ✓ **Data minimization:** AI systems must process only the data strictly necessary for their purposes.
     - ✓ **Automated decision-making and profiling (Article 22):** Individuals have the right not to be subject to decisions based solely on automated processing if these decisions significantly affect them (e.g., loan approvals).
     - ✓ **Transparency requirements:** Organizations must explain how AI-driven decisions are made, especially in contexts like credit scoring or fraud detection.
2. **Markets in Financial Instruments Directive II (MiFID II)**
   - → **Relevance to AI:** AI is increasingly used for algorithmic trading, portfolio management, and investment advisory services.
   - → **Key Impacts:**
     - ✓ **Algorithmic trading:** MiFID II mandates that algorithmic trading systems be monitored, tested, and risk-controlled to avoid market disruptions.
     - ✓ **Suitability and appropriateness:** AI-driven investment recommenda-

tions must meet the requirements for suitability and appropriateness tests under MiFID II.

- ✓ **Transparency:** AI-generated financial advice and trade executions must comply with pre- and post-trade transparency obligations.

3. Artificial Intelligence Act (AIA) (proposed)

- → **Relevance to AI:** The AIA is the EU's proposed regulatory framework for AI across sectors, including financial services.
- → **Key Impacts:**
  - ✓ **High-risk classification:** AI systems used in creditworthiness assessments or fraud detection are considered high-risk and will require compliance with stricter standards on transparency, accountability, and human oversight.
  - ✓ **Transparency obligations:** Financial institutions must disclose when AI is used and provide explanations of its functioning.
  - ✓ **Risk management:** Developers and users of AI must implement risk management systems to mitigate potential harms.

4. Payment Services Directive 2 (PSD2)

- → **Relevance to AI:** AI technologies are employed for fraud detection, risk assessment, and customer authentication in payment services.
- → **Key Impacts:**
  - ✓ **Strong Customer Authentication (SCA):** AI must ensure compliance with PSD2's requirements for secure user authentication while maintaining usability.
  - ✓ **Fraud detection and prevention:** AI-based systems must align with PSD2 mandates for monitoring and detecting unauthorized or fraudulent transactions.

5. Digital Operational Resilience Act (DORA)

- → **Relevance to AI:** AI is increasingly embedded in financial IT systems, creating potential vulnerabilities.
- → **Key Impacts:**
  - ✓ **Resilience of AI systems:** Financial institutions must ensure that AI systems used for critical operations are robust, secure, and capable of withstanding disruptions.
  - ✓ **Incident reporting:** Cyber incidents involving AI systems must be reported under DORA requirements.

6. Consumer Credit Directive (CCD) (revised proposal)

- → **Relevance to AI:** AI is used in credit scoring and loan approval processes.
- → **Key Impacts:**

- ✓ **Transparency:** Consumers must be informed about the role of AI in credit assessments.
- ✓ **Bias and fairness:** AI systems must avoid discrimination and bias in credit decisions.
- ✓ **Explainability:** Lenders must explain how AI-driven decisions are reached and provide recourse mechanisms for consumers.

7. **European Banking Authority (EBA) Guidelines**
   - → **Relevance to AI:** The EBA has issued guidelines on governance and risk management in the context of technological innovation.
   - → **Key Impacts:**
     - ✓ **Ethical AI:** Financial institutions must ensure AI is used ethically, avoiding discriminatory outcomes.
     - ✓ **Model risk management:** AI models must be validated, monitored, and tested to ensure accuracy and reliability.
     - ✓ **Accountability:** Institutions must have clear accountability frameworks for AI deployment.
8. **Anti-Money Laundering Directive (AMLD)**
   - → **Relevance to AI:** AI plays a critical role in detecting suspicious transactions and combating money laundering.
   - → **Key Impacts:**
     - ✓ **Enhanced due diligence:** AI-driven systems must comply with AMLD requirements for identifying and reporting suspicious activities.
     - ✓ **Transparency in risk assessments:** Financial institutions must explain and document AI-driven AML processes.
9. **Insurance Distribution Directive (IDD)**
   - → **Relevance to AI:** AI is used in underwriting and claims processing in the insurance sector.
   - → **Key Impacts:**
     - ✓ **Transparency:** Insurers must disclose the use of AI in pricing or claims decisions.
     - ✓ **Fair treatment of customers:** AI systems must not result in unfair discrimination in premium calculations or claim approvals.
10. **Capital Requirements Regulation (CRR) and Directive (CRD V)**
    - → **Relevance to AI:** AI is used for credit risk modeling, stress testing, and other regulatory capital calculations.
    - → **Key Impacts:**
      - ✓ **Model validation:** AI-based models must meet the stringent validation

requirements set by regulators.

✓ **Stress testing:** AI must enhance, not compromise, the reliability of stress-testing frameworks.

Emerging Areas of Focus

→ **Ethical AI frameworks:** The EU is increasingly emphasizing ethical considerations, requiring financial institutions to ensure fairness, accountability, and transparency in AI applications.

→ **Regulatory sandboxes:** Regulators are establishing sandboxes to encourage innovation while ensuring compliance with financial regulations.

By addressing these regulatory areas, financial institutions can navigate the complexities of AI integration in the EU's regulatory environment while ensuring compliance and fostering trust.

Gemini's response:

The use of AI in financial services is rapidly evolving, and its potential impact on EU financial regulations is significant. Here are some of the key regulations that may be impacted or potentially impacted:

Existing Regulations:

→ **General Data Protection Regulation (GDPR):** While not explicitly focused on AI, the GDPR's principles of data protection and privacy will still apply to the use of AI in financial services. This means that financial institutions using AI must ensure that they are processing personal data lawfully, fairly, and transparently, and that they have appropriate safeguards in place to protect data privacy.

→ **Markets in Financial Instruments Directive (MiFID II):** This directive governs the conduct of investment firms and the transparency of financial markets. While not directly addressing AI, it may be relevant to AI-powered trading algorithms and other automated trading systems.

→ **Capital Requirements Regulation (CRR):** This regulation sets out capital requirements for banks and investment firms. AI could potentially impact the calculation of risk exposures and capital requirements, particularly for models used to assess credit risk or market risk.

→ **Payment Services Directive (PSD2):** This directive governs payment services in the EU. AI could be used to enhance fraud detection and prevention systems, but it must comply with the security requirements and consumer protection provisions of PSD2.

New Regulations:

→ **AI Act:** This landmark legislation, currently under negotiation, will establish

a comprehensive regulatory framework for AI systems in the EU. It will likely have a significant impact on the use of AI in financial services, particularly for high-risk AI systems. The AI Act will likely address issues such as transparency, accountability, and human oversight in AI systems.

→ **Digital Operational Resilience Act (DORA):** This regulation aims to improve the operational resilience of financial institutions. It will likely require financial institutions to consider the risks associated with the use of AI, including the potential for AI-related failures or malfunctions.

It is important to note that the regulatory landscape for AI in financial services is still evolving. As AI technologies continue to advance, it is likely that new regulations will be introduced to address specific challenges and risks.

Additional Considerations:

→ **Model Risk Management:** Financial institutions using AI models must have robust model risk management frameworks in place to ensure the accuracy, reliability, and fairness of their models.

→ **Algorithmic Bias:** AI systems can be susceptible to bias, which can lead to unfair or discriminatory outcomes. Financial institutions must take steps to mitigate algorithmic bias in their AI systems.

→ **Cybersecurity:** The use of AI in financial services can increase the risk of cyberattacks. Financial institutions must have strong cybersecurity measures in place to protect their AI systems and sensitive data.

By staying informed about the latest regulatory developments and best practices, financial institutions can effectively harness the power of AI while mitigating the associated risks.

www.finance-watch.org

About Finance Watch

Finance Watch is an independently funded public interest association dedicated to making finance work for the good of society. Its mission is to strengthen the voice of society in the reform of financial regulation by conducting advocacy and presenting public interest arguments to lawmakers and the public. Finance Watch's members include consumer groups, housing associations, trade unions, NGOs, financial experts, academics and other civil society groups that collectively represent a large number of European citizens. Finance Watch's founding principles state that finance is essential for society in bringing capital to productive use in a transparent and sustainable manner, but that the legitimate pursuit of private interests by the financial industry should not be conducted to the detriment of society. For further information, see www.finance-watch.org.

Finance Watch
Rue des Colonies 56 box 3
1000 Brussels
T: + 32 (0)2 880 0430

www.finance-watch.org

Finance Watch