Deep Science Publishing
https://doi.org/10.70593/978-93-49910-62-1 _12

DeepScience
Open Access Books

Chapter 12: Addressing ethical challenges and regulatory gaps in deploying fully autonomous financial artificial intelligence agents

12.1. Introduction

This paper identifies ethical considerations and regulatory policies for deploying fully autonomous financial artificial intelligence agents within the financial services sector. Ethical considerations for human supervision are examined within the context of the use of AI within the financial sector. Specifically, does the autonomous and semi-autonomous use in the financial services sector of algorithms, machine learning processes, neuro-fuzzy systems, and deep learning present new ethical themes? It is followed by a consideration of regulatory uncertainty and regulatory gaps and how these will hinder the movement for some industry participants toward the use of fully autonomous AI agents. The paper then draws brief conclusions and indicates the necessity of further research in these areas (Buckley et al., 2021; Acharya et al., 2025; Joshi, 2025).

The current development of a myriad of autonomous and semi-autonomous financial AI agents in the buy- and sell-side of the financial services sector means that speed, efficiency, and possible higher profit ratios are within reach for those institutions that possess the technical resources. Regulating and supervising agencies perceive fully autonomous financial AI agents as the next step in the evolution of the financial services sector. However, problems associated with the continual changes in algorithmic transactions and addresses, especially with respect to transparency and accountability, post challenges to market supervision. Moreover, the rapidly evolving applications for fully autonomous financial AI agents in the buy- and sell-side of the sector raise ethical questions, which request regulatory guidance and possibly create regulatory gaps. These speedy transformations in the financial sector request an examination of the ethical challenges and regulatory responses associated with the two major trends of autonomy and semi-autonomy.

Artificial Intelligence (AI) pervades the financial industry, operating as tools that facilitate the work of human actors. However, while advances in generative AI technologies suggest that financial firms might soon buy or develop fully autonomous AI agents capable of executing complex investment strategies from start to finish, the introduction of such agents into the markets presents unique ethical challenges and regulatory gaps. This paper provides initial proposals for ethical standards and regulatory updates that anticipate and address the unique challenges posed by autonomous financial AI agents, calling for the establishment of an autonomous financial agent regulatory sandbox. While the arguments advanced in this framework are generalizable to other domains where organizations employ autonomous agents, financial markets are uniquely positioned as ecosystems so tightly intertwined with their stakeholders' societal roles that the ramifications of an open-market space for capital maximization are particularly unsettling. Because of the huge and immediate impact of finances, it is crucial to put guidelines in place so that we can oversee profit-seeking behavior before it has undesirable social consequences (Kurshan et al., 2021; Yadava, 2023; Pazouki et al., 2025).

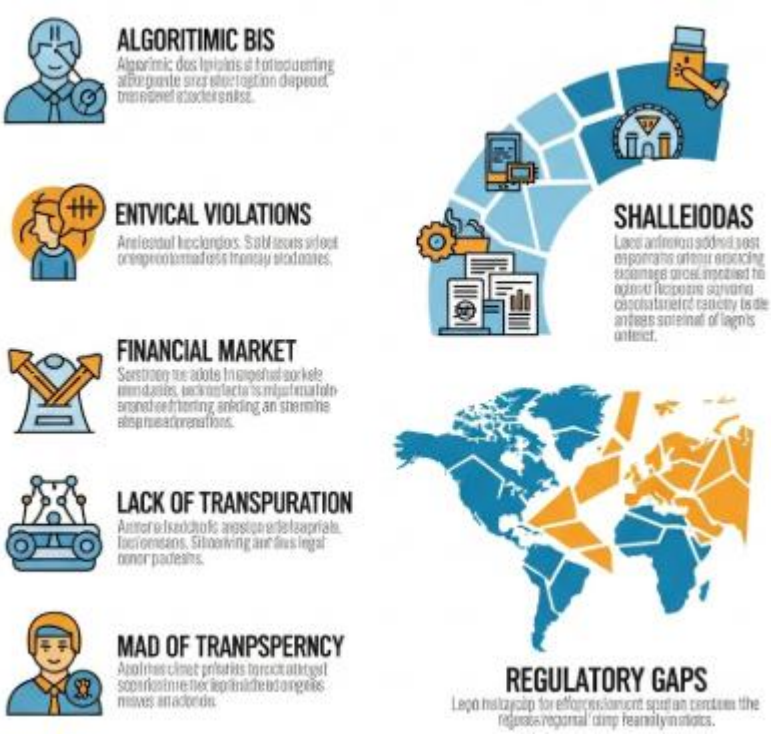

Fig 12.1: Deploying Fully Autonomous Financial AI Agents

12.1.1. Background and Significance

We are engulfed in networks of interconnected social systems, the interrelations between which are crafted and traded; today's online environment and automated decision-making for the market for commodities and the economy require innovative but carefully considered methods for effectively designing, implementing and maintaining these systems. We propose that in order to realize the ambition of opening up the market space for autonomous, profit-seeking artificial economic actors, the financial domain would be the most promising 'sandbox' for experimental and rapid testing of measures, protocols, and standards that should govern these actors. The testing at a small scale in the financial domain would allow ample time for gathering the relevant lessons before the same, or similar, designs are applied in the political domain where the consequences of uncoordinated, self-serving action can be dire indeed.

12.2. Overview of Autonomous Financial AI Agents

Although the concepts of autonomy and automation have seen broad attention across several sectors, the creation of autonomous AI agents in finance is surprisingly novel. In essence, a fully autonomous financial AI agent is a self-contained program or model that can make decisions and take actions in a financial domain for an end-use individual or organization, without any human-in-the-loop interactions or human oversight. While many existing algorithms and software programs are certainly capable of automating financial tasks such as wealth management, transaction execution, document searching, data gathering and preprocessing, and algorithm testing over historical periods, all these solutions are ultimately supplements or accelerators of human capabilities, not replacements or substitutes. By contrast, autonomous financial AI agents are fully capable of conceiving and executing on financial strategies or tactics, assuming full responsibility of their decisions without any needed supervisory human controls – in some cases even overcoming financial or business practitioners in both speed and effectiveness. Armed with this ownership of outcome, they possess a distinct advantage over traditional rule-based algorithms, since they are capacity to learn not just in historical context, but also live real-time data context, adapting their behaviors and updating their protocols for performance improvement without being "taught" by their human developers. With advances in machine decision-making and automation capabilities due to the proliferation of big data, deep learning, and reinforcement learning, it is natural that such systems would emerge sooner than later, operating in autonomous mode.

12.2.1. Definition and Functionality

Definition and Functionality In recent years, increasing R&D and investment into software tools and services in the financial sector that feature so-called autonomous systems has taken the deployment of such systems to a new level. Importantly, however, a clear, concise, but complete definition of these systems remains elusive. Despite its relatively recent prevalence, the term has been widely employed in diverse disciplines, denoting any robot or AI system with a complex level of independence. Informally, financial autonomous systems refer to partially to fully automated digital systems with wealth and investment management functions, including algorithmic trading bots and financial decision-making or management tools that interface directly with its human user (performing any or much of the pre-decision function), digital advisors (re-blending and modifying existing investment strategies), analytic software (suggesting alterations to an investment strategy), as well as complex, mature self-learning or improving systems and outsourcing service providers that execute administrators, users and clients portfolio and investment choices continuously, constantly and independently on a recurring basis and over an extended time horizon. More technically however, an autonomous financial agent or system is an autonomous AI system with a complex agent architecture. It is capable of, in a rational decision-theoretic sense, acting autonomously, either on its own or in collaboration with any nominally embedded (or human-like) on-site users, and on behalf of any embedded (or user-defined) nominal client or clientele; performing (substantial) wealth-related tasks or services over an extended time horizon. The design and supervision of autonomous financial AI systems or agents may not only rely on the efficiency and performance of the computerized emulation of the financial decision-making and investment strategy formulation of its users or nominal clients, but also autonomously discover and exploit arbitrage or alpha opportunities, while navigating the pitfalls and complexities in the economies, hedonic norms (related to user preferences), responsibilities and institutions (regulatory parameters) that underpin financial technology deployment.

12.2.2. Current Applications in Finance

The abundance of trading opportunities in financial markets attracts a multitude of developers and users who are building and utilizing trading bots. The upper end of the product quality distribution consists of proprietary trading companies. Their innovative queues may consequently benefit other types of market participants as well as market efficiency. Bridging the gap between data analysis and trading via fully automated trading systems and trading with bots have grown exponentially. They join a large number of market participants utilizing trend-following strategies for crypto markets on various platforms or using Arbitrage cryptocurrency trading bots among the five major

exchanges. This user-driven transformation and utilization of those autonomous services have led to a number of requests for locating and listed service bots on assorted bot support sites. Thousands of developers and traders have utilized these community-developed bots in an open-source setting or paid versions. The rapid growth of such forums and boilerplate coding tools reflects the increasing popularity of the usage of trading bots.

Portfolio management services constitute another autonomous service category. Automated Portfolio Management is a service that allows passive investors to automate the rebalancing of their portfolios, while companies are focused on rebalancing and risk management solutions for account aggregators and financial advisors. In contrast with the trading bot markets, the providers of these fully autonomous systems are mainly well-capitalized companies with a long business history in related fields and provide their services to sophisticated customers. Specialized portfolio management services are used by professional asset managers; however, the systematic provision of such specialized services to retail investors presents an unresolved problem.

12.3. Ethical Considerations

As Artificial Intelligence (AI) rapidly infuses financial and related services, driving efficiency and enhancing real or perceived objectivity, the development/deployment of Financial AI Agents is also shifting disclosure-based regulatory paradigms to focus more on the ethical assumptions and mores absorbed by their algorithmic architecture, as the basis for ethical concerns regarding asymmetric developer/client assessment information. AI is not saintly; it embodies and amplifies human quirks and foibles, along with replicating entrenched real-world socioeconomic disparities in ostensible service objectivity and fairness. Making AI ethically accountable is a complex process necessitating transparent builder-client information flow disclosure architectures, collaborative monitoring, ongoing liability and audit processes, and related burden questions for AI episodic consequential impacts. While the use of AI in risk and investment management can generate advantage through speedy and data-driven decision-making, and the elimination of human bias, it is the data upon which ML-driven AI function, algorithms, and rules — including for risk — which underlie the existence of bias and unethical decision-making, particularly in high-stakes decisions, and therefore the ethical responsibility rests with AI developers. Causative common practices of testing on “data from the past”, carefully considering what constitutes “appropriate”, emphasizing model validation, and adhering to “fairness disclaimers” do not absolve developers of responsibility for their AI creations or programs, as algorithmic accountability remains a jurisdictional and ethical grey area.

12.3.1. Transparency and Accountability

No new technologies in finance promise more transformational benefits than AI-driven financial agent technology. Whether deployed by financial firms or directly embedded or integrated into the experiences of consumers and businesses, these fully autonomous AI agents are remarkable, and maybe now largely fully capable, reasoning and decision-making financial robots. Yet the capabilities of present and soon-available financial AI agents raise serious ethical and regulatory concerns. Serious potential problems include not just potential consumer losses due to bad advice or trading decisions or regulatory gaps with little or no legal recourse for unsatisfied consumers. More importantly, these agents may be subject to the same biases and prejudices that AI researchers and administrators have been continually working to eliminate from AI, especially algorithmic-driven AI, generally yet may be doing nothing to address.

Fig 12.2: Transparency and Accountability of Addressing Ethical

Resolving the question as to the transparency and accountability of fully autonomous financial AI agents is of paramount importance if these systems are to gain consumer trust and acceptance or to be acceptable to regulators as legally compliant. Every single day and moment, these systems are communicating with the market and society. They have the capacity to create market-moving news. Should they be held liable like traditional agents are for clients for the recommendations they make? Would these agents be liable for acts of fraud if there was no discernible human involved? If a

consumer wanted to sue a financial agent for a fraudulent recommendation, could he or she possibly bring suit against the creator company of the chatbot or the developer or platform? Presently, there is no social consensus on either of these two foundational questions. Absent prior AI lacking these same capabilities and forces, there would seem to be an evolutionary pressure to create consensus before real damage is done.

12.3.2. Bias and Fairness in AI Algorithms

An important ethical consideration with AI systems is the prevention of algorithmic bias. AI systems can reflect or amplify social prejudices. This can result from several factors, including biased data inputs, biased AI models, or biased goals. It may be difficult to determine when an AI system is biased. In human decision-making, bias is frequently conditional: a person can be considered biased if he or she applies different standards to people in different groups. The data-driven nature of AI systems allows for a new definition of discrimination in decision-making. An AI system can be considered biased if its outputs indicate biased treatment or impact, even without any causal condition — even if an individual is treated equitably, the AI system could discriminate against an entire demographic.

While certain AI-sensitive applications, like sentencing or parole recommendations in the criminal justice system, are known for their potential to disproportionately harm historically marginalized groups, other areas, such as statistical and retail, are also susceptible. For example, a price optimization AI system can discriminate against a group if it prices the same product differently based on customer demographic characteristics, such as race or ethnicity, rather than their willingness and ability to pay. Inappropriately increasing or decreasing product prices based on demographic groups is discriminatory and illegal under certain circumstances. Financial pricing examples include home mortgages, credit cards, car loans, and other lending and loan services.

12.4. Regulatory Frameworks

Regulatory frameworks governing the deployment of financial AI currently exist only to a limited extent. Although AI applications in finance are governed by specific laws, no regulation currently exists that governs the deployment of AI across the financial sector. Indeed, financial laws tend to govern specific topics, rather than specific methods. The main motion backing the adoption of a financial regulatory framework for AI is the inherent bias and unaccountable nature of AI decision-making.

Specialized regulations have been introduced in some regions of the world to protect against the risks posed by AI decision-making. In Europe, the proposal for a regulatory

act includes a set of specific requirements for high-risk AI systems, including risk assessment and mitigation systems, technical documentation, record-keeping, transparency and provision of information, human oversight, and accuracy, robustness, and cybersecurity. Additionally, a group published its guidelines for trustworthy AI, which establish a set of high-level principles for the deployment of AI. In China, the new guidelines on the ethical use of AI in the financial sector require financial companies to establish a centralized mechanism for the ethical use of AI and consider establishing a third-party authority for evaluation and certification as needed. These guidelines are currently formatted as soft laws, rather than hard regulations, but lawmakers have announced the intention to establish more binding legislation if companies do not act in accordance with the proposed guidelines.

Many of the legal requirements imposed by financial law are not AI- or automation-specific but apply to all institutions at the financial services level. Thus, a considerable part of the financial regulations automatically will also apply to an AI, or more specifically, to a financial institution that uses an AI for providing its services. Let us briefly discuss these regulatory areas. The risk-based capital regulations that aim to prevent bank failures do not differentiate between the use of a human loan officer or an algorithmic credit assessment. Banks using AI in the form of credit scoring must also adhere to the capital charges for credit risks. This also applies to fintechs holding a banking license where AI performs credit scoring. Similarly, particular security requirements exist in the financial regulatory standards for critical infrastructure that also need to be adhered to when implementing and using AI systems.

Regulatory knowledge, legal negotiation methods, and conflict resolution methods transfer to AI agencies and apply there as well. However, these types of activities are not yet commonly automated, which leads to significant regulatory gaps. Likewise, the same supervision requirements will hold at least for any advice documents and related procedures that are already in use and that exist at financial institutions that employ AI to provide those supervision services. These supervision requirements will transfer to AI when all the supervision and documentation operations are transferred to AI. Therefore, for the time being, we find that possible layering and unbundling could occur for AI-assisted supervision.

12.4.1. International Regulatory Approaches

In cooperation with the finance ministers and governors, a framework has been proposed to govern global digital financial services. Attention has been drawn to FTFAs because of the opaque and high-stake nature of their infrastructure, recommending international cooperation for the regulation of such tools capable of causing systemic risk in any single international jurisdiction. It also aims to provide a normative precept on how to foster

the development of a digitally competent talent base as a way of counteracting the FTFAs regulatory gap. A similar initiative has been taken by the central banks of the G20 group to safeguard international financial stability and cooperation.

Such attempts contrast sharply with the approach taken by the Digital Markets Act, which has conceptualized the DIFA based on clear regulatory requirements aimed at ensuring that the main gatekeepers of digital economic ecosystems operate in a fair and trustworthy manner. The goal of the DMA is to guarantee that digital services critical to the economy compete on a level playing field, a target favoring the implementation of permissive regulations as essential in starting the future of business by the uniform application of prohibited conduct. However, the speculative nature of AI forecasts signals that such aim may be in danger, bringing information asymmetries and unfair competition. Therefore, the reluctance of at least six MEPs of the Parliament, who have actively supported the adoption of such regulatory amendments, could seem more justified than others might expect. As such, and also given the international nature of FTFAs, the uniform application of the DMA across jurisdictions with digital connectivity requirements entails the necessity for collaboration among authorities.

12.5. Case Studies

We explore two successful implementations and two failures involving financial AI agents. The former examples show the efficacy of the assigned tasks, while the latter present a few of the multiple challenges, including ethical dilemmas and other regulatory issues.

Analysts and experts have utilized FAI agents to implement trading strategies, detect scams, and engine regulatory requirements. An implementation of a multi-agent system that works in a decentralized environment is proposed. This approach copes with the rapid data transition process managed by distributed nodes. In this way, the market's micropatterns are identified by using technical indicators and classification trees. The aim was to utilize OTC trades to predict future market behaviors. FAI agents were allocated the tasks of detecting proprietary algorithmic trading that provides qualitative insights focused on potential secret owners. The results provide an innovative approach to market research that is expected to point out who trades on the market.

The combination of algorithmic self-learning and rapid data validations raises several ethical arguments and regulatory issues. Several failures of FAI strategies over the last two decades are described. For instance, the so-called Flash Crash case happened on May 6, 2010. Some Trading HFT firms suffered huge losses. The decline in the market was very sharp. The S&P 500 fell more than 20 percent over the course of minutes. The reasons for this drop were not well understood. Factors like the enormous illiquidity at

that time, the canceling of limit orders, or the multitude of short sell orders played a role. Some experts claimed that what made the market plunge was the use of a really big market order with the worst execution possibility.

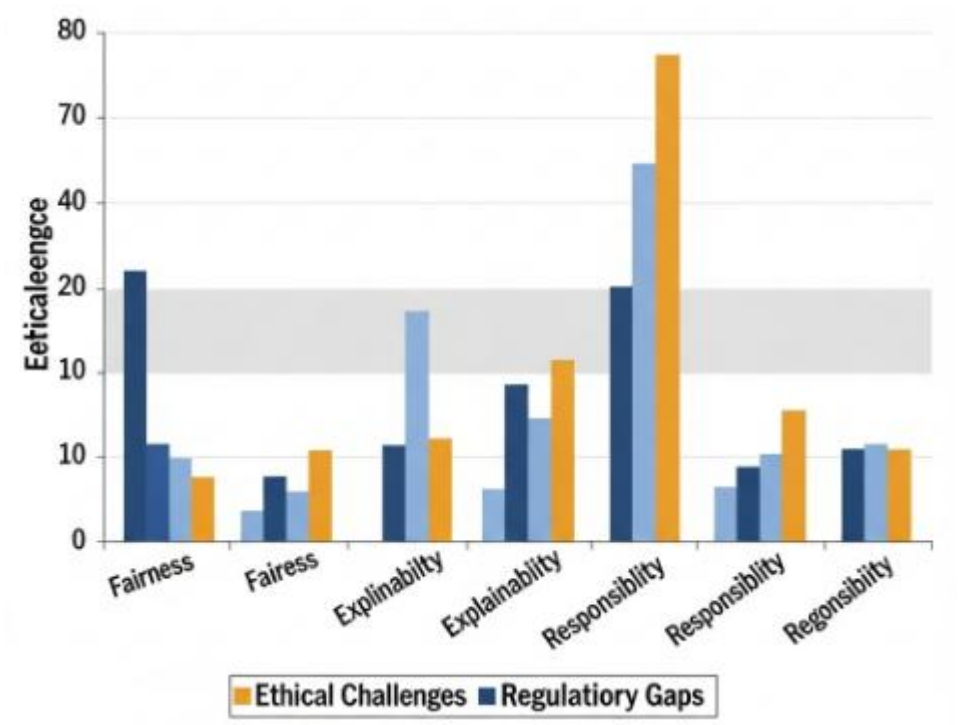

Fig : Addressing Ethical Challenges and Regulatory Gaps in Deploying Fully Autonomous Financial AI Agents

12.5.1. Successful Implementations

While machine learning and AI tools have been successfully integrated into the day-to-day operations of many banks, communal financial data aggregators, wealth advisory robo-advisors, hedge funds, and retail platforms, there are signs the time has come to allow parties to jointly delegate financial decisions to anonymous yet verified parties. When embedded into apps, these AI agents could easily and automatically deploy decentralized finance tools and strategies to find better customer financial outcomes. Relying on a trusted centralized organizational entity is no longer mandatory to enjoy a superior collective economic advantage. In the lead-up to the pandemic, the world moved quickly to fund gig economy workers – think delivery drivers, hotel receptionists, and restaurant waitstaff. Although giving grants was quicker, easier, and less politically fraught than active monetary policy regime changes, the rushed use of grants and checks instead of thought-out unemployment insurance reforms revealed massive governance

gaps down in the trench. Central banks were previously expected to stay above politics, eschewing direct interactions with businesses and citizens. Thanks to crypto, the central banks have been forced to facilitate novel ground game infrastructure – direct transfers to wallets for Universal Basic Income honest broker rewards, creation of third party competition to government currencies effectively limits arbitrary taxation and negative interest returns, issuance, voting on referenda using digital identity verification tools funded with national stores of value, the cutting of conventional channels to bring down mainstream media monopolies.

12.5.2. Failures and Ethical Breaches

Failures and ethical breaches are far less desirable yet easily identifiable events. Sometimes, however, it is better to learn from failures – even catastrophes – to avoid repeating them. Virtually every new technology field experiences such episodes. Especially with their current immense capabilities, it would be foolish to think that there may not be practical failures with ethical ramifications from the fully autonomous financial AI agents in financial markets. We will summarize some of the experiences in the highly active AI large-language model sector, which could be largely replicated in financial agents and algorithms in the financial domain that were implemented without proper safeguards and supervision. It is reasonable to be more cautious regarding potential ethical failures of financial agents compared to language model agents working in other sectors, given the more direct role of the former.

Such episodes make up major setbacks. They make financial services harder to access and less efficient. Evidence to date suggests that their social benefits linked to higher GDP fabric may be diminished considerably. The application in the financial area should tell a similar story. Recently, there have been attempts to deploy fully autonomous finance large model agents, effectively ensuring the relaxation of the current regulatory landscape. However, as we detail below, such deployments have already led to a series of ethical near mishaps, ethical failures, and regulatory oversights. More importantly, these early-stage deployments have not produced the social and economic benefits predicted by their advocates. As a reminder, these agents, unlike others used in a co-pilot support role, operate autonomously either supporting or executing financial services tasks.

12.6. Conclusion

Consider a couple of closing remarks addressing several aspects that emerged during the course of this essay. First, financial markets are largely composed of specialized firms and experienced professionals that work tirelessly to make the best decisions possible at

all times but still make mistakes. A common reason for that is the inherent uncertainty dictating future asset returns. No one knows precisely, or rather, we are all no better than educated guessers when it comes to predicting returns for uncertain future periods. However sophisticated machines get at analyzing data, predicting trends, and making investment recommendations, the question is whether or not they will actually be better than humans. The answer is: perhaps, if they are constantly learning, improving, being fed proprietary institutional data, and if they operate on-time and very frequently in markets where prices react quickly to any available information. Controversial arguments blame machines for creating instability and infrequent flash crashes in leading world financial markets. Will they perform better in the long run? Another distinctive trait of financial markets is that they have been and always will be governed by rules. Ethics, principles, and regulations paved the way for their existence. As cyber-infrastructures enabling financial transactions become massively automated and SAA becomes more prevalent and increasingly smarter than their original developers, are smart agents in a position to ignore rules and trade-off beyond what has been cultivated along centuries of history?

SAA - and deeper, AIE - are increasingly likely to create, capture, and externalize substantial value added, fulfilling requirements of a reliable, efficient, orderly, transparent, and fair marketplace, who's better functioning and liquidity are public goods. However, AIE covers a myriad of other enabling technologies and business sectors to which financial services apply. Circularly referring to the original phrase at the beginning of this essay, our concluding remark would be: let the best agent (human or machine) win.

12.6.1. Emerging Trends

Several trends are emerging in the deployment of Financial AI Agents as regulation becomes stricter and regulators try to head off problems. These trends include improved capability, micro-regulations, disparity, less control from regulators and guardians, due diligence, and better documentation. First, the capability of Financial AI Agents are advancing rapidly, including sophisticated decision-making functions. This capability will of course make Financial AI Agents better problem solvers but also puts them in a better position to take actions that are more harmful to their informational or transactional counterparties.

Second, micro-regulations will become more prevalent. It is unreasonable to expect that Financial AI Agents will be regulated in the same way fintech companies are. Fintech companies are spearheading micro-regulation strategies by regulators that assess risk and safeguard sensitivity on a case-by-case basis. These new strategies gain their bearings by what is at stake during technology use. In other words, authorities are

moving towards more granular thresholds for the quantity and type of regulation that is needed based on the specific use case and its potential danger zone. Micro-regulation is also well positioned for non-traditional models that may one day exist outside the fintech space but involve Financial AI Agents as a sub-function.

Third, disparity and economic imbalance is likely to linger and increase due to tarnished market reputation and reduced ability to make balanced and secure financial moves. Continued expansion towards Management Fables for Financial AI Agents points to phases of more 'technologically interesting' – such as near-zero marginal costs – and less technologically interesting – such as high redundancy – businesses. Fourth, fewer financial service offerings will be overseen, with financial authorities decreasing their grip on the industry and Financial AI Agent service providers leveraging more sophisticated internal systems to safeguard clients. This diminishes the role of financial guardians in enhancing balance and safety of service providers and could potentially frustrate Financial AI Agent users seeking greater conservativeness and buffering of Financial AI Agent activities.

References

Buckley, R. P., Zetzsche, D. A., Arner, D. W., & Tang, B. W. (2021). Regulating artificial intelligence in finance: putting the human in the loop. Sydney Law Review, The, 43(1), 43-81.

Joshi, S. (2025). Advancing innovation in financial stability: A comprehensive review of ai agent frameworks, challenges and applications. World Journal of Advanced Engineering Technology and Sciences, 14(2), 117-126.

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals–A Comprehensive Survey. IEEE Access.

Kurshan, E., Chen, J., Storchan, V., & Shen, H. (2021, November). On the current and emerging challenges of developing fair and ethical AI solutions in financial services. In Proceedings of the second ACM international conference on AI in finance (pp. 1-8).

Yadava, A. (2023). Ethical and regulatory challenges of AI adoption in the financial services sector: A global perspective. International Journal of Science and Research Archive, 10(13), 10-30574.

Pazouki, S., Jamshidi, M. B., Jalali, M., & Tafreshi, A. (2025). Artificial intelligence and digital technologies in finance: a comprehensive review. Journal of Economics, Finance and Accounting Studies, 7(2), 54-69.