# Bounded autonomy: AI agents reshape finance and raise urgent governance stakes



Autonomous AI agents are rapidly evolving from digital assistants into independent actors capable of executing complex financial transactions—introducing a governance challenge as profound as the technology itself. Kathryn McCall, Chief Legal and Compliance Officer at Trustly, is among those urging a structured, cautious approach to deployment, grounded in transparency, accountability and ethical safeguards.

McCall champions the principle of "bounded autonomy": clear limits on what AI agents can do, layered governance structures, and human oversight at key decision points. In practice, this means permitting AI to initiate tasks like invoice creation, but requiring human approval for critical actions such as payments. The goal is to manage the growing risks around financial privacy, compliance and security as AI agents assume greater operational control.

Unlike traditional financial software, AI agents often behave unpredictably. Their non-deterministic nature introduces vulnerabilities including prompt injection, adversarial attacks and data leaks. McCall highlights the potential “blast radius” of AI failure—how much damage can result when things go wrong. To mitigate this, she proposes infrastructure safeguards like sandboxing, isolated containers, time-limited access and emergency kill switches, particularly for high-stakes actions such as cross-border payments.

The regulatory landscape remains patchy. While comprehensive AI-specific laws are still emerging, existing regulations on data privacy, anti-money laundering and payment processing already apply. McCall warns that companies must not view the current regulatory gap as a green light to innovate without responsibility. GDPR and financial accountability standards must be reinterpreted for the age of AI agents.

The broader industry is undergoing dramatic change. AI is improving customer service, accelerating credit decisions and supporting fraud detection—but also enabling more sophisticated financial crime. Industry experts stress the need for explainability, ensuring AI decisions are transparent and understandable to humans from the outset.

Risks extend beyond finance. Legal liability, cybersecurity, bias and physical or financial harm all increase as AI agents gain autonomy. Experts recommend full risk assessments, clearly defined accountability structures and responsible AI principles built into corporate practice. PwC and others advocate detailed operational guidelines and strong stakeholder engagement to uphold trust in AI-driven systems.

This shift elevates the corporate legal function. McCall argues that legal officers must lead the effort to map AI behaviour to legal exposure and mitigation frameworks, ensuring ethics and recoverability remain core values. Automation may streamline processes, but it must never replace human responsibility.

Despite the risks, autonomous AI offers significant opportunities for customer experience, efficiency and compliance. With rigorous, principle-led governance and proactive oversight, the UK can lead in responsible AI innovation. Embedding bounded autonomy, explainability and regulatory foresight into system design is a critical step towards building trusted, resilient AI in financial services.

Created by [Amplify](https://www.hbmadvisory.com/amplify): AI-augmented, human-curated content.

## Bibliography

1. <https://www.pymnts.com/news/artificial-intelligence/2025/growth-of-ai-agents-put-corporate-controls-to-the-test/> - Please view link - unable to able to access data
2. <https://www.pymnts.com/news/artificial-intelligence/2025/growth-of-ai-agents-put-corporate-controls-to-the-test/> - This article discusses the rapid evolution of autonomous AI agents in the financial services sector, highlighting their transition from simple assistants to autonomous decision-makers capable of executing financial transactions. Kathryn McCall of Trustly emphasizes the need for 'bounded autonomy,' advocating for internal controls, human oversight, and sandbox environments to mitigate risks and ensure accountability. The piece underscores the importance of companies proactively implementing transparency, auditability, and ethical guardrails in AI systems, especially in the absence of clear regulations.
3. <https://www.ft.com/content/da0f4df3-72bd-481d-a3c1-222a406e7ba2> - This Financial Times article explores the transformative impact of generative AI on the financial services industry, noting its applications in enhancing customer service through advanced chatbots and automating lending processes. It also highlights AI's role in investment, aiding fund managers in data analysis, and in combating financial crime by expediting anti-money laundering efforts. However, the piece also addresses concerns about AI enabling more sophisticated financial crimes and the need for robust governance to prevent unintended consequences.
4. <https://www.reuters.com/legal/transactional/legal-transparency-ai-finance-facing-accountability-dilemma-digital-decision-2024-03-01/> - This Reuters article examines the challenges posed by AI algorithms in the financial sector, particularly regarding transparency and accountability. It discusses the 'Black Box' nature of AI, which raises concerns about explainability in critical areas like credit scoring. The piece also covers regulatory challenges, including data governance, privacy, and compliance with laws such as GDPR, and notes the EU's pending AI Act aimed at regulating AI risks and promoting ethical practices.
5. <https://www.reuters.com/legal/legalindustry/ai-agents-greater-capabilities-enhanced-risks-2025-04-22/> - This Reuters article delves into the dual aspects of AI agents in various industries, highlighting their advanced capabilities and the heightened risks they introduce. It discusses concerns such as physical harm, privacy violations, biased outputs, legal infractions, and cybersecurity threats. The piece emphasizes the importance of robust AI governance frameworks, thorough risk assessments, and clear legal responsibilities to mitigate these risks effectively.
6. <https://www.forbes.com/sites/bernardmarr/2025/03/11/forget-chatgpt-why-ai-agents-are-the-real-game-changer-for-financial-services/> - This Forbes article highlights the transformative potential of AI agents in the financial services industry, noting their ability to function as autonomous analysts and investment advisors. It discusses how AI agents can enhance risk assessment, credit scoring, and regulatory compliance by autonomously analyzing data and adapting strategies. The piece also touches on the importance of responsible AI practices to ensure these agents operate ethically and effectively.
7. <https://www.pwc.com/us/en/tech-effect/ai-analytics/responsible-ai-agents.html> - This PwC article discusses the integration of AI agents into financial services, emphasizing the need for responsible AI practices. It outlines strategies for aligning AI agent deployment with responsible AI principles, including setting clear operational guidelines, defining roles and responsibilities, and ensuring stakeholder engagement. The piece also highlights the importance of embedding responsible AI practices into reporting ecosystems to enhance integrity, auditability, and quality in financial reporting.