AIFS0061

Written evidence submitted by Dr Swati Sachan,
University of Liverpool-Management School

AI Collaboration to Counteract Flaws in High-Stakes Financial Decisions

Summary of Evidence:

- Q1: AI-driven financial systems face challenges on decision opacity, sparse or unreliable data, and security vulnerabilities. Additionally, the demands for future scalability require a robust computational infrastructure for equitable outcomes in high-stakes financial domains.
- Q2: AI outperforms human experts in repetitive (rule-based) tasks but falters when faced with ambiguous data and novel scenarios, whereas financial experts excel with contextual reasoning and ethical judgment.
- Q3: AI's "bias" is not intrinsic but a reflection of flawed human inputs. Human cognitive limitations: redundant data, biases (predictable deviations), and judgment "noise" (unexplained variability) contaminate AI systems by embedding inconsistent human decisions into training data.
- Q4: The potential of Decentralised Finance (DeFi), powered by blockchain's immutability (tamper-proof data permanence) and transparency (publicly auditable), to address the scarcity of reliable financial data and accountability in AI-driven systems.

This evidence note presents the benefits and risks to consumers arising from increased integration of AI decision-support systems for high-stake financial decisions. It synthesizes scientific findings from our research on explainable AI algorithms for transparent decision-making and Decentralised Financial (DeFi) technologies to address security- and privacy-related issues, tested across FinTech companies, insurance providers, and community-driven microenterprises. Through real-world applications, we identify critical challenges on data scarcity, privacy vulnerabilities, and the progressive contamination of AI systems over time due to reliance on redundant data and inconsistent judgments by human experts, which perpetuate historical biases embedded in past policy and regulatory practices. Following is our response by four core questions:

Question 1: What challenges hinder the ethical deployment of AI decision-support systems for autonomous high-stakes financial decision-making? How does it affect vulnerable consumer populations?

The UK's financial services are already at the forefront of AI-driven digital innovation, which has fundamentally reshaped consumer engagement and operational efficiency across banking, investment, and insurance markets. However, the unprecedented speed of AI advancement in the post-generative AI era has introduced critical risks to data integrity, AI model reliability, and governance frameworks, which threaten the sector's stability and equitable service delivery (Bank of England, 2023).

The UK government's pro-innovation approach is committed to regulating AI for responsible and safe deployment (UK Government, 2024). The Bank of England, along with the Prudential Regulation

Authority (PRA) and the Financial Conduct Authority (FCA), supports these efforts by actively monitoring safe AI innovation within financial services and issuing guidelines that prioritize secure AI functionality, transparent and fair decisions, and legal accountability for AI governance (Bank of England, 2024).

We identified challenges based on our scientific research and understanding of AI risks, especially after the unprecedented speed of AI advancement in the post-generative AI era:

(i) Decision Transparency and Explainability: AI algorithms deployed in financial services process sensitive datasets containing demographic attributes, behavioral metrics, private financial history, and sociocultural characteristics, which raises the risk of biased and discriminatory outcomes. For high-stake financial decisions such as loan approvals or liability insurance outcomes, regulatory frameworks on GDPR mandate explainability (the ability to articulate the rationale behind decisions) and interpretability (the technical capacity to trace how and what pattern model has found in the input data to produce output).

Evidence from Study 1 on Interpretable AI: In collaboration with a UK-based FinTech lending firm, we developed an inherently interpretable hybrid AI model that can be trained by data and can incorporate human judgments to process first-charge and second-charge mortgage loans (Sachan, Yang, Xu, Benavides, & Li, 2020) (Sachan, S., 2022). Given the safety-critical nature of loan underwriting in high-impact financial contexts, the model architecture prioritized mathematical traceability to ensure transparent and accurate decision justifications to financially viable individuals with prior credit deficiencies. These individuals face systemic exclusion from traditional banks due to rigid scoring models. We found that end-to-end transparency from input data to algorithmic processing to final output generation provides a granular understanding of the following:

(a) How much data support (number of samples) is available for a particular type of loan application to give high-confident or low-confident decisions? For instance, joint mortgage applications by civil partners exhibited ambiguous outcomes due to insufficient samples (underrepresentation) reflecting shared credit histories with conflicting financial behaviors.

(b) Which attributes in the dataset were most important in deriving an outcome of a loan application? For instance, most rejected loan applications had more than three payday loans in the past two years, bankruptcy in the past five years, and lack of crucial credit history.

(c) Stakeholders can understand the accountability of adverse decisions due to complete transparency. For instance, the model's explanation in some cases revealed that a loan denial disproportionately weighted outdated bankruptcy records (a decade old) over recent fiscal responsibility.

Evidence from Study 2 on LLM as Black-Box Model: Unstructured data, such as financial reports, customer correspondence, and transaction voice logs, cannot be effectively processed by inherently interpretable AI models, which lack the complexity to analyse high-dimensional, non-linear patterns (discussed in Study 1). These data types demand mathematically complex algorithms such as deep neural networks (DNNs) or transformer architectures, which excel in tasks such as image classification, speech recognition, and semantic text analysis, but they are "black-box" in nature. Their internal decision-making process is opaque and interpreting the correct reasoning behind the outcome is not impossible but difficult.

Manual human processing of large amounts of text data in repositories of financial institutions is time-consuming (Bholat, Hansen, Santos, & Schonhardt-Bailey, 2015). Large Language Models (LLMs), built on transformer architectures powering generative AI tools such as ChatGPT, DeepSeek, and Grok, have demonstrated unprecedented capabilities in processing vast textual data and approximating human-like reasoning (Sachan, S; Dezem, V; Fickett, D, 2024). However, analysing sensitive financial data on third-party LLM platforms introduces significant risks of data breaches and leakage of proprietary or personal information with an unstable explanation of outcomes due to external system control.

To address these challenges, our research investigated the responsible deployment of LLMs within financial institutions (Sachan, S.; Miller, T.; Nguyen, N. M., 2025).

(a) Human-in-the-Loop Governance: We developed and trained a localized LLM in a secure, in-house environment to eliminate external vulnerabilities. A human-in-the-loop framework was implemented to monitor LLM performance continuously by quantifying prediction uncertainty and confidence levels. Low-confidence outputs, such as generated text or labels, were flagged and redirected for manual review. We proposed a technique on lexicon robustness to explain a decision from LLM.

(b) Local LLM: Open source LLMs, Bert-large-uncased, Mistral, LLaMA2, and LLaMA3 models were deployed locally to process financial reports of small businesses. The robustness of LLM's explanation was evaluated by underwriters; approximately 81% were accurate and contextually coherent, while the remaining 19% of low-confidence cases required refinement. For instance, a loan denial citing "insufficient collateral" was revised after a human review revealed undetected assets in the report. It demonstrates the necessity of human-AI collaboration in high-stakes decisions beyond technical accuracy.

(c) Computing Infrastructure: We want to point out that local LLM deployment has the potential for transparent outcomes and prevention of data leakage by keeping models and proprietary financial data secure in a private environment. It could balance innovation with accountability and security. However, hosting localized LLMs requires large computation resources such as

powerful GPUs (Graphical Processing Units). Public-private partnerships could subsidize computational infrastructure for smaller institutions to democratize access to secure AI tools.

(ii) Scarcity of Reliable Financial Data: In another research project, we observed a depletion in the accuracy of AI-driven financial decisions due to a lack of high-quality, human-generated training data (Sachan S. , Almaghrabi, Yang, & Xu, 2024). Research pointed to the collapse of AI models due to recursive training with synthetically generated data (Shumailov, et al., 2024). The financial ecosystems are dynamic due to policy shifts and evolving consumer behaviour. Therefore, historical financial data must be continuously updated to reflect these changes; failure to do so can lead to adversarial decisions. For instance, preliminary audits of our first-version AI underwriting system revealed decision inaccuracies traceable to obsolete lending criteria embedded within static training datasets. This case study will be expanded in Questions 2 and 3 of our response.

(iii) Data and AI System Security: Data security risks extend beyond leakage to third-party platforms; financial institutions face targeted cyberattacks that threaten sensitive data and disrupt centralized AI systems. These systems are susceptible to adversarial attacks, where malicious actors manipulate the inner workings of the model (parameters) or inject distorted data, leading to flawed decisions. Such breaches can cause significant financial losses, harm larger demographics, and trigger serious legal consequences (Sachan, Fickett, Kyaw, & Purkayastha, 2023) (Sachan, S.; Liu, X., 2024). Our research proposes safeguarding AI systems with decentralised technologies, such as blockchain, discussed further in Question 4 of our response.

Question 2: Under what conditions do AI systems outperform human financial experts in high-stakes decision-making?

The automation of repetitive, rule-based controlled tasks through AI has drastically reduced human error and operational costs. However, this efficiency comes with a critical limitation: AI performs well in environments with clear rules and predictable outcomes but struggles when faced with ambiguity, incomplete information (data), or novel scenarios demanding contextual reasoning (Sachan S. , Almaghrabi, Yang, & Xu, 2021). This gap emerged in our research at a UK-based FinTech firm, where we developed and evaluated a hybrid explainable AI decision-support system for financial underwriting (Sachan, Yang, Xu, Benavides, & Li, 2020) (Sachan S. , Almaghrabi, Yang, & Xu, 2024). We found that complex underwriting decisions, such as the evaluation of applicants with irregular financial histories or risk profiles, cannot be processed by the transparent AI system but heuristic expertise: domain-specific knowledge professionals cultivate through years of experience. Human underwriters consistently demonstrated the ability to resolve ambiguities in unstructured data (e.g., vague income documentation), adapt to regulatory shifts, and balance ethical priorities that rigid algorithmic frameworks could not replicate.

Question 3: How do redundant data, cognitive biases, and inconsistencies in human judgment contribute to the gradual contamination of AI decision-support systems?

Human cognition is superior in understanding and analysing contextual ambiguous information. However, empirical evidence challenges the presumed reliability of human judgment: cognitive biases, expertise disparities, and subjective interpretations generate systemic "noise" manifested as inconsistent decisions by the same expert (intra-individual variability) or divergent outcomes across experts (inter-individual variability) in identical high-stakes contexts (Kahneman, Rosenfield, & Blaser, 2016). This raises a governance dilemma: can inherently "noisy" human judgments serve as a valid benchmark for training and validating autonomous systems in high-stake domains such as finance, insurance, law, and healthcare? Nobel laureate Daniel Kahneman distinguished bias as predictable deviations from ground truth from "noise" as unexplained variability in judgments (Kahneman, Sibony, & Sunstein, 2021). Marginalization by discriminatory outcomes arises when systems codify flawed assumptions from unrepresented and contaminated datasets. Redundant data, cognitive biases, and inconsistencies in human judgment collectively degrade AI systems through a self-reinforcing cycle of algorithmic contamination.

We explored the process of degradation of AI systems by contaminated training data in an investigative study implemented in a FinTech firm (Sachan S. , Almaghrabi, Yang, & Xu, 2024). An explainable AI system based on the evidential reasoning-explainer (ER-X) algorithm was trained on a dataset of 5,700 mortgage loan applications. The dataset was aggregated from credit bureaus, fraud intelligence, and digital loan applications, comprised 26.8% rejected and 73.2% funded cases. The attributes on property valuation, credit score, and affordability status had the highest reliability score which implies that data for these lending rules had caliber to point strongly toward a decision. However, a noise audit exposed how human judgment variability introduces flaws into such systems.

Four underwriters, two senior and two juniors, evaluated three lending profile types processed by the system: 16 clear-cut decline criteria, 36 recurrently funded profiles, and 42 recurrently rejected profiles within the firm. Results revealed that senior underwriters consistently diverged from juniors. Junior underwriters leaned on simplified algorithmic outputs and rigid rules (e.g., auto-rejecting applicants with over six telecom arrears) due to their limited experience. Senior underwriters manually adjusted 37% of decisions based on their deep domain knowledge of policy shifts and crisis management, especially for borrowers with weak credit histories.

Question 4: How can stakeholders collaboratively address data scarcity in AI decision-support systems to ensure equitable financial inclusion? In what ways can innovations in Decentralised Finance (DeFi) serve as a solution to the data scarcity and privacy challenges?

DeFi is a financial ecosystem built on Distributed Ledger Technologies (DLT), such as blockchain. Blockchain stores data in a linear chain of cryptographically linked blocks. DeFi innovations are based on two properties of DTL: transparency due to public audibility (any network participant to independently verify data integrity and source authenticity) and immutability (permanence and

inalterability of recorded data once validated). These properties diverge from centralized ledger systems, where opaque, institutionally controlled governance mechanisms restrict public access. DeFi uses these features to let people lend, invest, share information, and trade digital assets, such as cryptocurrencies, non-fungible tokens (NFTs, such as digital art), and other digitized resources(data and documents) in a trustless environment through automatic agreements written in code called smart contracts.

DeFi for Inclusive Entrepreneurship (Sachan, Fickett, Kyaw, & Purkayastha, 2023)**:** We harnessed blockchain technology to address the financial barriers faced by small businesses in securing funding. In research with community-based microfinance institutions, we recognized that some of these barriers are due to data scarcity in banks and credit bureaus, which typically do not serve many small businesses due to their insufficient or non-existent credit histories. The fragmented data infrastructure and opaque financial ecosystems limit the ability to generate the verifiable records necessary for fair credit assessments. We designed a blockchain-based architecture for financial institutions to aggregate knowledge and data in a secure and joint consent-driven environment. The system complies with data protection regulations EU GDPR on the "right to erasure." The data is shared in an encrypted and anonymized format to safeguard privacy. Additionally, all consents related to data sharing or termination are permanently recorded on the blockchain, providing a transparent and accountable audit trail. The aggregated data is then utilized to power a transparent AI-driven decision-support system.

The initiative aggregated anonymized financial data from four institutions to understand credit default of individuals with mid-range credit scores (500–650), fewer than 2 defaults in 36 months, gender parity (49% women, 51% men), and 68% newer businesses (<5 years). Analysis showed a low default rate (1.8% annualized), suggesting undervaluation of community financial behaviors. Women-owned enterprises had 18% higher revenue retention, indicating opportunities for gender-responsive lending.

Blockchain to Manage Accountability for Generative AI Outcomes in Insurance (Sachan, S.; Liu, X., 2024)**:** In another research, we proposed and tested (an insurance law firm) a solution to find an equilibrium between responsible usage and control of human professionals over content produced by Generative AI through real-time automated audits. It investigates the potential of Generative AI in drafting correspondence for pre-litigation decisions on liability insurance. Results suggest that blockchain's tamper-proof record-keeping enhances transparency in AI-generated content to foster trust among insurers, claimants, and regulators on the ethical use of AI.

Acknowledgements and Contacts:

This evidence notes is part of an ongoing impact case at the University of Liverpool: “Explainable AI and Blockchain Solutions for High-Stake Decisions: Capital Access to Underserved Communities” (https://www.liverpool.ac.uk/management/research/impact/explainable-ai-and-blockchain-solutions/).

AIFS0061

This response has been prepared by University of Liverpool-Management School, Financial Technology Research Group (Dr. Swati Sachan).

April 2025

References

Bank of England. (2023). *FS2/23 – Artificial Intelligence and Machine Learning.* Feedback statement 2/23.

Bank of England. (2024). *Artificial intelligence in UK financial services.*

Bholat, D., Hansen, S., Santos, P., & Schonhardt-Bailey, C. (2015). Text mining for central banks: handbook. *Centre for Central Banking Studies Handbook*, 1-19.

Fickett, D. (2023). *How Enterprise Development Can Help Heal America's Divisions.* Forbes Business Council.

Kahneman, D., Rosenfield, A., & Blaser, T. (2016). Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making. *Harvard Business Review*.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment.* Hachette UK: Little, Brown and Company.

Sachan, S. (2022). Fintech lending decisions: an interpretable knowledge-base system for retail and commercial loans. *Cham: Springer International Publishing*, (pp. 128-140).

Sachan, S., Almaghrabi, F., Yang, J. B., & Xu, D. L. (2021). Evidential reasoning for preprocessing uncertain categorical data for trustworthy decisions: An application on healthcare and finance. *Expert Systems with Applications*, 115597.

Sachan, S., Almaghrabi, F., Yang, J. B., & Xu, D. L. (2024). Human-AI collaboration to mitigate decision noise in financial underwriting: A study on FinTech innovation in a lending firm. *International Review of Financial Analysis*, 103149.

Sachan, S., Fickett, D. S., Kyaw, N. E., & Purkayastha, R. S. (2023). A Blockchain framework in compliance with data protection law to manage and integrate human knowledge by fuzzy cognitive maps: small business loans. *IEEE International Conference on Blockchain and Cryptocurrency*, 1-4.

Sachan, S., Yang, J. B., Xu, D. L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, 113100.

Sachan, S.; Liu, X. (2024). Blockchain-based auditing of legal decisions supported by explainable AI and generative AI tools. *Engineering Applications of Artificial Intelligence*, 129, 107666.

Sachan, S.; Miller, T.; Nguyen, N. M. (2025). Responsible LLM Deployment for High-Stake Decisions by Decentralized Technologies and Human-AI Interactions. *IEEE International Conference on Human Machine Systems (Accepted Upcoming May 2025)* (pp. 1-6). Abu Dhabi: IEEE.

Sachan, S; Dezem, V; Fickett, D. (2024). Blockchain for Ethical and Transparent Generative AI Utilization by Banking and Finance Lawyers. *World Conference on Explainable Artificial Intelligence* (pp. 319-333). Malta: Springer Nature Switzerland.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 755-759.

UK Government. (2024). *A pro-innovation approach to AI regulation: government response to consultation.* Department for Science, Innovation, & Technology.