

A Study on Human-Centric Explainable Ai Auditing Systems for Financial Inclusion, Digital Empowerment, and Trust-Building in Emerging Economies

Ganesh P^{1*}, Avinash Kumar Namdeo²

1 Research Scholar, Department of Computer Science Engineering, Lingaya's Vidyapeeth,
Faridabad, India

ganesh05@gmail.com

2 Associate Professor, Department of Computer Science Engineering, Lingaya's Vidyapeeth,
Faridabad, India

Abstract: This paper investigates human-centred XAI resolutions for financial inclusiveness, digital enablers and responsibility and confidence in AI-enabled finance in emerging markets. It examines how AI aids financial audit across credit, fraud, regulation, audit, compliance, digital payment services and risk classification; and it enables exclusionary hazards of unequal data, unaccountable tools, inscrutable docs and opaque management; ultimately aiming to develop an Human-Centred XAI Trust Structure for contextually-appropriate audibility and contestability in identifying if financial audit is technically reliable, intelligible, challengeable, socially inclusive and institutionally accountable. It employs qualitative secondary data-based conceptual research using an integrative literature review, thematic investigation, qualitative content analysis, stakeholder examination, case-based comparative analysis, risk-control activity mapping and conceptual synthesis; drawing evidence from open access academic journals, policy papers, financial innovation policies, AI ethical codes, fintech management policies, emerging markets exemplars. Its essentials signify that AI-enabled financial audit provides dual accountability due to the double nature of financial activity and its automated audit; and technical transparency even cannot generate peoples' understanding, capability, correction and inquiry of the automated decisions. Financial digital exclusion can arise from incomplete and unrepresentative data, proxy-based bias, uneven error rate, opaque docs, natural language translation, human fragility, disintegrated accountability lines and weak grievance response. The context-sensitive human-centred trust structure weights up contextual perception, data and modeller integrity, human-centred explication, audibility, participation, inclusive management and empowerment outcomes. It documents Stakeholder-centred explication, responsible decisions, Trust calibration and Guide to contextual implementation for responsible AI in digital finance.

Keywords: Human-Centric XAI; Financial Auditing; Financial Inclusion; Digital Empowerment; Trustworthy AI

1. INTRODUCTION

AI is "computers doing tasks that would usually require human intelligence", and ML is a form of AI whereby systems learn from data how to make predictions and decisions. ML based AI is already being more frequently used in financial markets and regulatory markets to evaluate creditworthiness, approve loans, prevent fraud, find anomalies, to monitor anti-money-laundering and to undertake auditing of transactions, produce financial statements, evaluate insurance application, scrutinize customers risk profile, develop regtech and supotech.

Accordingly, there are two types of relevant auditing. Financial auditing AI involves AI use to review transactions, accounts, banks, borrowers and other financial activities, while auditing of AI involves AI application to processes, models, decision making, explanations, bias and discrimination, governance and social impacts.

Interpretable AI or XAI models: Help to enable effective understanding of individual results and model behaviour. - Explainability: provides explainability by illustrating the 'knowledge' behind a model, while transparency addresses accountability through transparency of data, misspecification, processes, limitations and institutional responsibility; - Algorithmic auditing: audits computational and information system design, information contained within the information system, algorithmic process design, data used by the information system and governance of the information system; - Human-centred AI: addresses the design of effective and human rights respecting AI; - Financial inclusion: brings tangible access to and use of financial services; - Digital empowerment: empowers effective understanding and control; - Digital trust: gives conviction and confidence; - Contestability: enables review and contestation; - Recourse: provides capacity for challenge and redress; - Inclusive AI governance: provides training and harmonised decision-making.

II. Conceptual framework and literature review In developing countries, mobile banking, mobile money, digital lending, digital identification, alternative-credit data, fintech platforms and digital public infrastructure compress the physical gaps, transaction costs, documentation requirements, and service-delivery lag. – but account or application ownership do not necessarily mean financial inclusion: for disadvantaged groups, connectivity gaps, limited devices, digital skill gaps, language gaps, credit-scarcity gaps, consent gaps and complaint gaps may still hinder access and use (Klapper et al., 2025). The Global Findex 2025, comprising nationally representative surveys of ca. 145 000 adults in 141 economies, investigated the correlations between use of financial services, mobile phone ownership, internet access and digital safety; and remaining gender and poverty gaps (ibid.). Inclusive finance means access, use, literacy, agency, protection, redress, trust and inclusive outcomes.

Standard AI-enabled audit systems aim to optimise accuracy, speed, risk detection and operational efficiency - but can produce rational decisions about those affected persons' behaviour, which cannot be understood, verified or challenged. Technical opacity makes nuanced interactions in complex models opaque; social opacity marginalises language, literacy

and cultural conditions; institutional opacity imputes responsibility to the entire financial system, fintech vendors, external auditors and regulators; procedural opacity places notice, correction and appeal out of stakeholders' reach; and distributional opacity reflects sampling bias, proxy variables and inferior digital access for disadvantaged groups. The real problem is not that AI is difficult to explain, but that the reasons it offers for decisions will be insufficient for affected persons' real agency, redress or protection.

Recent developments have emphasised explanation instead of stakeholders' understanding, prioritising explanation through lifecycle auditing, human review and independent accountability. The NIST AI Risk Management Framework linked explainability to transparency and accountability, privacy, reliability and combating harmful bias through the functions of govern, map, measure and manage the AI system (Tabassi, 2023). The revised OECD AI Principles again mention 'understandable information' and traceability, human agency and means for stakeholders adversely impacted by AI to protest AI outcomes (Organisation for Economic Co-operation and Development [OECD], 2024). Human-centric explanation has thus become a governability matter rather than a technical feature.

The literature under review is based on human-centred AI, explainability (XAI), algorithmic accountability, procedural justice, capability-based digital empowerment, financial inclusion, and trust in institutions. Human-centred AI considered human rights and human limitations first when conceptualising systems. XAI delineated technical and user interface tools explaining decisions; algorithmic accountability required documentation, traceability, monitoring and remedial measures; procedural justice called for reasons for which stakeholders would consider both financial, and communication, review and appeal, processes just; and digital empowerment meant the real capacity to understand, act, decide, challenge and choose, including the option to calibrate or withhold trust, confident that the system performs competently and in accordance with establish institutional accountability.

Raji et al. (2020) developed a comprehensive internal algorithmic auditing framework that integrated responsible AI documentation and responsibility with lifecycle review. Erneviien and Kabainskas (2024) found 138 relevant researches for the main financial purposes of XAI including credit management, financial prediction and fraud detection. Kim et al. (2024) found that only a technical solution for explainability had not yet sufficiently managed meaningfulness, user-understandability and human-AI interaction. Maathuis et al. (2026)

found that from the users' (and other professional stakeholders') view in decision-making, financial explanation formats needed to be human-centred. Previous researches had not integrated the dual auditing approach, financial inclusion, digital empowerment, stakeholder engagement, contestability, institutional legitimacy, emerging-economy circumstances and SDG-minded governance into one comprehensive framework.

The present research seeks to produce a Human-Centric XAI Trust Framework for evaluating AI-enabled financial auditing in emerging economies. A qualitative, secondary-data driven, conceptually oriented research design was chosen employing integrative literature review, qualitative content analysis, thematic analysis, stakeholder analysis, risk-control map, comparative case analysis and framework development. Considered open access academic and institutional sources only, published 2018-2026, purposively selected to be relevant, reliable, transparent about their methodology and credible; but academic, institutional, paid and paid-indexed, duplicated, promotional, retracted and bibliographically non-verifiable sources excluded. Findings were descriptively mapped, textually coded, thematically categorised, cross-analysed for stakeholders and comparative emerging-economy examples, synthesised for the framework, and formulated into operational propositions.

2. DUAL-LAYER AI AUDITING AND EXCLUSION RISKS IN EMERGING FINANCIAL ECOSYSTEMS

AI-Assisted Financial Auditing and Regulatory Decision Chains

Financial audit-supporting technologies To aid financial audit, artificial intelligence applications may help conduct automated transaction testing, implement continuous auditing and fraud detection and prevention, provide credit-risk classification and conduct customer due diligence, enable anti-money-laundering monitoring, validate insurance claims, conduct tax analytics and supervise digital lending. Firms use machine-learning algorithms to process large volumes of both unstructured and structured data, uncover anomalous activity patterns and forecast default risk and to generate alerts that may not otherwise be visible through manual processes. Technologies to support regulatory compliance and supervisory tools would assist firms and supervisors respectively with the monitoring of risks to financial stability stemming from institutions, market activity and financial infrastructure. AI can strengthen know-your-customer processes and detect suspicious activity patterns by aggregating information on transactional relationships (Aldasoro et al., 2024). The Financial Action Task

Force (2021) further noted technologies would enable a faster, more precise and better use of anti-money-laundering and counterterrorist-financing measures.

The complete decision chain can be expressed as:

Data collection → data preparation → model processing → risk score or audit flag → institutional interpretation → financial or regulatory decision → stakeholder communication → review or appeal.

It should be able to be tracked down at every stage. Data can be accountable to a bank, a third-party, a public authority or a fintech service. It can be constructed by in-house teams or by technology providers. The output can be assessed and used by an auditor, a compliance function or a regulator - but not necessarily held to account: The vast majority of decisions are approved by senior managers, but it's the borrower, policyholder, taxpayer, recipient of a benefit or small business that bears the risk of the automated decision Human oversight is also not necessarily accountable when the reviewers do not have the necessary skills, authority, time or documentation.

Opacity, Bias, and Accountability Gaps

Dual layer audits need to examine the behavior of the AI under review and the tools through which it generated that behavior (Goldblum, et al, 2023). Risks can pertain to data (such as gaps, spurious correlations, digital footprint inequality, training data sample bias, missing informal data, consent and data drift), model-level risks (black box predictions, overfitting, unstable feature importance explanations, error rate differences across subgroups, subgroup overfitting), leading to an incorrect classification of applicants from underrepresented groups despite system performance being acceptable overall (ibid). Technical explanation/justification/decision-related risks encompass generic rationale, technical jargon, and insufficiently detailed explanations that do not 'disarm' malicious actors; organizational issues such as reliance on third-party vendors, incomplete audit records, opaque accountability relationships, uninformed staff, and conflict between commercial sensitivities and public accountability; and governance-related concerns in circumstances where authorities lack the technical skills, standardized methodologies for audits are lacking, data protection standards are insufficient, enforcement powers are limited, and grievance mechanisms are weak or non-existent (Prez-Cruz, et al, 2025). Prez-Cruz, et al (2025) observed that post-hoc explanation methods may be approximate, unstable and unreliable and that auditors may not be able to

evaluate models without access to the full ecosystem of emerging technologies. Interdependent features of exemplability and auditability include the possibility that an overly simplified explanation may be provided for a prediction even if regulators cannot verify the entire process or that extensive technical detail exists even if it does not provide an explanation or narrative for the individual affected.

Differential Impact on Vulnerable and Marginalized

Digital populations Not everyone can expect to enjoy equal access to AI-powered credit risk assessments. For women entrepreneurs, informal workers, borrowers living in rural and lower income households, minorities of language, migrants, older age, disabled people, and financial literacy gaps may have a smaller digital footprint. Beyond that measurable alternative data for those with limited or no traditional credit history, and to measure other social and "intrusive" data points (Mhlanga, 2021), might let vulnerable populations be served by alternative data.

A common exclusion pathway is:

Limited credit history → dependence on alternative data → opaque risk inference → adverse decision → inadequate explanation → inability to correct or appeal → continuing exclusion.

Women entrepreneurs might be particularly adversely affected by differences in device ownership, lower levels of overall asset ownership, limited collateral, reduced documentation of business records and by being previously absent from bank data. Some characteristics such as location, store characteristics, shopping behaviour, device usage, the frequency of transactions, and others could, intentionally or unintentionally, uphold gender inequality if the gender as a variable is excluded from the model. Kelly and Mirpourian (2021) identified that algorithmic lending can unlock credit for women but could also embed old discrimination into the data.

Table 2.1: AI Auditing Exclusion Risk Map

Data or System Condition	AI-Generated Decision	Mainly Affected Groups	Immediate Harm	Long-Term Consequence
Incomplete formal credit records	High-risk classification	Informal workers and micro-enterprises	Loan rejection or higher interest	Continued dependence on informal credit
Gendered or proxy-based data	Lower credit score	Women entrepreneurs	Smaller loans or denial	Reinforcement of gender financing gaps
Poor identity matching	Fraud or compliance alert	Migrants and mobile workers	Account restriction	Loss of access to formal finance
Inaccessible explanation	Unchallenged adverse decision	Rural, low-literacy and disabled users	Inability to correct errors	Reduced trust and digital withdrawal
Weak human review	Automatic confirmation of model output	All vulnerable stakeholders	Procedural unfairness	Institutionalised financial exclusion

The map of risks suggests exclusion is unlikely to result from one technical error, but a chain of differential data scenarios, automated categories, unavailable explanations and inadequate remedial actions. When auditing for exclusion, over time, both model performance but also sub-group effects, quality of explanations, human audits and financial impact must be considered.

3. HUMAN-CENTRIC EXPLAINABILITY, PARTICIPATION, AND DIGITAL EMPOWERMENT

Stakeholder-Centred Explanations and Usefulness of Explanations

Explainability from a human perspective recognises the divergence of explanatory requirements for different stakeholders in AI-enhanced financial decision-making. Financial citizens and consumers require simple, local-language, decision-shaping explanations that include inaccurate data and enable appropriate, multi-directional (contested) explanations of a decision e.g. an applicant for finance might prefer to know the identifiable, correctable qualities of her finances that resulted in a rejection, rather than the workings of an AI model. It was found that explanation needs can be generalised as "why this? " (Liao et al. 2020). Auditors require technical explanations, e.g. model documentation, data provenance, feature-level, validation metrics, subgroup performance, case logs and versions.

Regulators require system-wide explanations, e.g. accountability, legal compliance and impact assessments, institutional responsibility, fairness testing, vendor involvement, memory size, model and remediation documentation. Developers require model debugging, error analysis, feature interactions, invariance testing, concept drift and adaptation signals. Civil-society groups require aggregated data, e.g. on social-economic disparities, social impact, complaints and countermeasures. Thus, explanation type varies with purpose and stakeholders. Feature-importance explanations may suffice for auditors; counterfactual or local explanations, for consumers. System-wide explanations may be needed for regulators; visualisation and rule-based explanations, for accessibility. The following characteristics are necessary for an explanation to be effective: the facts must be correct, the explanation must be easy to understand, relevant, practical to implement, available at the time it is needed, consistent across cases, respect privacy, be transparent, be readily available, and transparently communicate the level of uncertainty in any given case In addition, Ehsan et al. (2021) observed that explanations need to be social and organizational in context.

Participatory Auditing, Contestability, and Human Oversight

Customer involvement has to be throughout the project, not just at the end. Setting problems, choosing data, characterising risks, validating models, designing explanations, monitoring and updating models, systems and decisions should all involve customers, auditors, regulators, developers, citizens and the victims of the system. Stakeholder workshops, consumer panels,

user-interface testing, civil-society audits, participatory impact assessments, and open government audits can all help expose risks that experts can't see.

Contestability A decision must be contestable. Contestability which assesses how feasible it is to contest a decision through the use of an AI. A decision is contestable if the individual is made aware of the use of AI and if they are provided with an accessible reason, if they can correct inaccurate data, if they can submit additional evidence, if they are supervised by a human, and if they receive an appropriate remedy within reasonable amount of time. Form of access to a human who will make sure that the algorithms have been accurately fi of attribute and criteria that are relevant to concerns" (Moreira et al.2025).

There are three methods of including human review: Human-in-the-loop, where humans are involved in every decision; Human-on-the-loop, in which a human operator monitors the system (and can step in if necessary); and Human-in-command, whereby humans are responsible for setting the goals of the system and when it is invoked and withdrawn. Human review alone may be little more than check-box compliance procedures if those reviewing the model predictions have insufficient technical expertise, resources, authority or evidence to inform their decisions The automation bias could even cause reviewers to blindly trust the AI product.

Digital Empowerment and Trust-Building Mechanisms

Having digital access and skills to navigate digital financial systems is called digital empowerment. Digital empowerment involves informed consent, data agency, access to help, the ability to seek redress and to be protected from harm, and the ability to intervene in or challenge the output of automated decision-making. Trust should therefore be measured on various layers of trust: trust in the system output, the financial institution, the auditor and regulator, the digital infrastructure and grievance process.

More explanation does not necessarily lead to more trust. An explanation might expose uncertainty, unfairness or lack of transparency or poor model performance. What we want is such calibrated trust that reliance is consistent with accuracy and transparency, as well as effective oversight and redress. Jacovi et al. (2021) identified warranted and unwarranted trust; warranted is well-founded trust based on actual system trustworthiness, while unwarranted is trust that is unfounded.

Table 3.1: Stakeholder–Explanation–Empowerment Matrix

Stakeholder	Decision Context	Information Need	Preferred Explanation	Required Support	Empowerment Outcome	Trust Implication
Financial customer	Credit approval or rejection	Reasons, incorrect data and improvement options	Plain-language and counterfactual explanation	Human assistance and appeal channel	Ability to correct or challenge decisions	Greater calibrated trust
Auditor	Model and transaction auditing	Data lineage, validation and error evidence	Feature analysis, logs and technical documentation	Full system access	Independent verification	Evidence-based institutional trust
Regulator	Compliance and consumer protection	System-level risk, fairness and accountability	Global explanation and impact reports	Enforcement and inspection powers	Effective regulatory oversight	Greater public confidence
Developer	Model design and monitoring	Error patterns, drift and feature interactions	Diagnostic and local-global explanations	User and domain-expert feedback	Improved system performance	Increased technical reliability
Civil-society	Group-level	Disparities	Aggregate and	Access to independent	Collective advocacy	Improved

organisati on	financial impact	complaints and corrective action	publicly understand able disclosure	ent audit findings	and participatio n	institutio nal legitimac y
------------------	---------------------	---	--	-----------------------	--------------------------	-------------------------------------

The matrix shows that explanations empower only if their form is right for the role of the stakeholder, the decision that must be made and their ability to act on it. Customers want explanations that are simple and actionable; auditors and regulators, want verifiable evidence to help them explain and justify decisions. Trust is built not by using the same language with every stakeholder but by giving each stakeholder explanations, support and intervention opportunities that matter to them.

4. DEVELOPMENT OF THE HUMAN-CENTRIC XAI TRUST FRAMEWORK FOR INCLUSIVE FINANCIAL AUDITING

Architecture of the Proposed Framework

Implementation of Human-Centric XAI Trust Framework The designed framework provides a multidimensional approach for assessing whether AI-supported financial auditing mechanisms are technically safe, socially inclusive, explainable and institutionally accountable. The framework's seven dimensions are integrally related because there can be damage at any stage of the machine-learning lifecycle: at data collection, during the development of models, in their evaluation and deployment (Suresh & Guttag, 2021). Version 3, September 2023.

4 Human-Centric XAI Trust Framework The first dimension, Context and Inclusion Assessment, looks at the target population, financial service, social disadvantage, digital access, financial literacy, language, connectivity, environment, and possible exclusion risks.

This assessment ensures that auditing mechanisms do not ignore the realities of women entrepreneurs, rural borrowers, informal workers and digitally marginalised communities. The second dimension, Data and Model Integrity, considers data quality, adequacy, provenance, consent, precision, resilience, proxy indicators, subgroup performance, model drift and uncertainty. The third dimension, Human-Centred Explainability, considers explanation,

including explanation for different stakeholders, comprehensibility, response capability, faithfulness, actionability, fair access, temporal perspective, provision in local language, human support when automated explanation is unavailable. The fourth dimension, Auditability and Traceability, considers decision logs, source of data, model records, post-hoc evidence, logs of incidents, agent-to-agent reporting. The fifth dimension, Participation, Contestability and Recourse, considers stakeholders' involvement, transparency of AI, remedy, human auditability, recourse, redress, stakeholders' feedback and right to explanation.

The sixth dimension, Inclusive and Institutional Governance, considers regulation, protection of financial consumers, third-party provider liability, multi-stakeholder audit, civil society participation, liability for harm. These dimensions correspond to widely accepted principles concerning human agency, technical robustness and safety, data governance, transparency and accountability (High-Level Expert Group on Artificial Intelligence, 2020). The seventh dimension, Trust, Empowerment, Inclusion and SDG Outcomes, considers human understanding, process fairness, data agency, access to finance, trust in public institutions, gender fairness and economic participation. UNESCO (2021) also recognized human rights, fairness, transparency, accountability, inclusiveness and multi-stakeholder governance as the overarching principles for human-centred AI.

Operational Indicators and Conceptual Propositions

Performance metrics for the framework design are the following: Technical metrics: accuracy of explanation, stability of explanation, correctness of explanation, robustness of explanation, subgroup mistakes of explanation, drift of explanation, completeness of data available, and diversity of explanation data. Human-centred metrics: explainability knowing, cognitive burden of explanation, relevance of explanation, actionability of explanation, accessibility of explanation, explanation of explanation accessibility in local languages, explainability assurance. Audit metrics: explanation's traceability, documentation of explanation, explanation validation, frequency of explanation review, mispredicting explanations detected, frequency of correcting explanations, correction-action rate.

Participation metrics: explanation's human critique, appeals of explanation accessibility, explanation review response time, explanation correction rate to mispredicting explanation, explanation appeals' rate of quality, explanation complaints' resolution rate.

Inclusion metrics: explanation acceptance gaps, gaps of explanation of rural and women populations, access by disadvantaged groups and explanation's participation in formal financial system. Governance metrics: governance common sense, explanation' compliance to rules, explanation responsibility intelligence, explanation openness to information, explanation's founding legitimacy. The hypotheses of the framework are: P1: Explicability to the people increased via explainability accuracy; P2: Explicability to the people increased via explicability transparency; P3: Explicability to the people increased via explicability knowledge; P4: People extralegal power; P5: Guardian extralegal power; P6: Power for explicability increases through explicability actionability; P7: Power for explicability increases through participancy; P8: Power for explicability increases through contestability; P9: Digital empowerment and access increases through digital explainability and accountability; P10: Accountability increases through explainability audit; P11: Financial inclusion increases through inclusive explainability and governance.

Emerging-Economy Implementation and SDG Alignment

Implementation should be appropriate to institutional capability. Low-capacity settings should focus on plain language, community mediators, local-language support, accessible complaints, and interpretable models. Implementation of building regulatory systems should include impact assessments, auditor training, subgroup testing, vendor discloser and incident reporting. More advanced digital states could implement continuous monitoring, interoperable audit logs, independent access, transparency registries, and participatory regulatory sandboxes. This framework supports SDG 1 on financial resilience, SDG 5 on gender equitable financial inclusion, SDG 8 on entrepreneurship and broad participation, SDG 9 on responsible digital infrastructure, SDG 10 on reduced inequality, and SDG 16 on accountable institutions (see Table 1). These findings demonstrate the spirit of the broader development agenda of the 2030 Agenda (United Nations, 2015).

5. INTEGRATED FINDINGS AND CONCEPTUAL SYNTHESIS

Principal Findings Emerging from the Analysis

The combined research showed that the audit of financial risks in an AI-related process faces two accountability issues. The use of AI in an audit was often applied to scan financial transactions for fraud, classify financial risks, observe regulatory compliance and contribute to financial decision-making. As an end in itself, the very AI systems used in the audits should

also be internally audited on their data sources, model behaviour, explanations, fairness, risk decisions and social impact. The scope of the audit could no longer be restricted to just financial documents.

Results showed that a technical explanation, on its own, was inadequate if affected stakeholders were unable to comprehend, contextualise and challenge an artificial intelligence (AI)-enabled decision. An explanation might be mathematically accurate, but socially ineffective if communicated in difficult words, if it was not produced in local languages, or if it was not paired with available solutions.

Financial exclusion was observed across multiple levels of interconnected factors. Data-level exclusion can occur due to incomplete, out-dated datasets, or models which are unintelligible (black-box). Model-level disadvantages can arise through use of generic explanations, limited human oversight, shared (lack of) organisational accountability, or limited regulations. Gender and other protected groups may be indirectly impacted via proxy variables such as location, use of the digital platform, job roles, type of device used, and transaction history.

An analysis of the types of explanation needed by specific groups was also carried out: financial customers need simple explanations and actions; auditors require convincing evidence, documentation and traceability; regulators require evidence at the system level of system fairness, control and accountability; developers need diagnostics, error descriptions and descriptions of the states, behaviour and performance of models; civil-society groups need evidence of grouping dynamics, bias and remediation. Human oversight was only possible if reviewers had the capacity and expertise, time and access to access to the necessary technical detail. Enabling participation and contestability implied that explanation extended beyond transparent presentation and disclosure to enable effective agency and empowerment through the capacity to modify data, present evidence, to call upon human oversight and, if necessary, to appeal. For trust to be established it must be appropriately calibrated in relation to evidence of system quality and acknowledged limitations, while strong institutional characteristics must be in place. The capacities of individuals (digital and financial literacy, language, and connectivity) also influence the effectiveness of XAI processes for agency and empowerment.

Theoretical, Technical, Auditing, Policy, and Development Contributions

The study's main theoretical contribution was the composition of a theoretical model including explainable artificial intelligence, algorithmic auditing, procedural justice, digital

empowerment, financial inclusion and institutional trust. The technical contribution was to extend the evaluation of XAI explanations beyond accuracy and explanation fidelity by evaluating comprehension, accessibility, actionability, subgroup effects and feedback of stakeholders. The contribution of the audit platform was the two-tier strategy. It integrated financial AI auditing supported by AI with independent auditing of AI. The technical contribution was to bring the assessment of the explanatory content of XAI explanations in addition to accuracy and explanation fidelity also to assess comprehension, usability, actionability, subgroup analysis and feedback of stakeholders. The contribution of the audit platform was the hybrid two tier strategy. It combined financial AI auditing (mobedev), enabled by AI with independent auditing of AI (one world platform). The policy contribution was the framework for social inclusion and accountability indicators and the assessment of the quality of AI governance by regulators and financial institutions, respectively. The development contribution was the responsible AI governance approach to be adopted for the socio-economically inclusive digital participation for the sdg rederacy in nascent market economies. policy contribution was the provision of social inclusion and accountability indicators and an assessment of quality of the governance by regulators and financial institutions, respectively. The development contribution was the of responsible AI governance to promot equitable digital participation, reduce exclusions and SDG oriented fiancial inclusion in emergnet economics.

Consolidated Human-Centric XAI Trust Framework and Outcome Mapping

Table 5.1: Human-Centric XAI Trust Framework for Inclusive Financial Auditing in Emerging Economies

Framework Dimension	Core Assessment Question	Main Mechanisms	Suggested Indicators	Expected Outcome	SDG Connection
Context and Inclusion Assessment	Who may benefit or be excluded?	Vulnerability mapping, contextual assessment and stakeholder identification	Gender, income, location, literacy, connectivity and access gaps	Context-sensitive system design	SDGs 1, 5 and 10

Data and Model Integrity	Is the system technically reliable and fair?	Data validation, proxy testing, subgroup evaluation and drift monitoring	Accuracy, error gaps, data quality, robustness and stability	Reduced discriminatory risk	SDGs 9 and 10
Human-Centred Explainability	Can each stakeholder understand and use the explanation?	Plain language, visualisation, counterfactuals, reason codes and local-language interfaces	Comprehension, fidelity, actionability, accessibility and timeliness	Meaningful understanding	SDGs 5, 9 and 10
Auditability and Traceability	Can the complete decision process be independently examined?	Audit logs, documentation, version records and independent validation	Traceability, documentation completeness, review frequency and corrective action	Institutional accountability	SDG 16
Participation and Contestability	Can affected stakeholders participate and challenge decisions?	Consultation, human review, appeal, data correction and	Appeal access, response time, correction rate and	Agency and procedural fairness	SDGs 10 and 16

		grievance mechanisms	participation level		
Inclusive Governance	Are responsibilities and oversight arrangements clear?	Regulatory supervision, vendor accountability, multidisciplinary auditing and public disclosure	Responsibility clarity, compliance, incident reporting and transparency	Legitimate and responsible governance	SDGs 9 and 16
Trust, Empowerment and Inclusion Outcomes	Does the system improve equitable participation?	Calibrated trust, financial capability, access support and institutional safeguards	Trust, service access, gender gap, rural access and successful recourse	Financial inclusion and digital empowerment	SDGs 1, 5, 8 and 10

The study's main theoretical contribution was the composition of a theoretical model including explainable artificial intelligence, algorithmic auditing, procedural justice, digital empowerment, financial inclusion and institutional trust. The technical contribution was to extend the evaluation of XAI explanations beyond accuracy and explanation fidelity by evaluating comprehension, accessibility, actionability, subgroup effects and feedback of stakeholders.

The contribution of the audit platform was the two-tier strategy. It integrated financial AI auditing supported by AI with independent auditing of AI. The policy contribution was the provision of social inclusion and accountability indicators and an assessment of quality of the governance by regulators and financial institutions, respectively. The development contribution was the of responsible AI governance to promote equitable digital participation, reduce exclusions and SDG oriented financial inclusion in emergent economies.

6. CONCLUSION AND RECOMMENDATIONS

Conclusion

This paper examined human-centred XAI for financial auditing, financial oversight, financial inclusion, digital empowerment and public trust in emerging economies. The research established that the human-centred use of AI enabled financial auditing to improve the quality, scalability, accuracy and efficiency of transaction monitoring, fraud detection, credit assessment, auditing, compliance and risk management - but not necessarily in a manner that was seen as fair, explainable, inclusive or trustworthy. The paper illustrated that stakeholder-sensitive explanations promote less financially excluded citizens as customers have the means to interrogate, contest and seek redress for non-transparent automated decision-making. The paper illustrated that stakeholder-sensitive explanations enhance transparency by more contextually and conversationally informed decision-making avoiding a one-size-fits-all approach to transparency about algorithm operations that fails to deliver counter-objective transparency for customers, auditors, regulators, data scientists and civil society. A calibration of trust was observed to be rooted in inclusive stakeholder participation, human oversight and contestability; pluralistic perspectives; accessible human-readable information; and responsible public regulation. The Human-Centric XAI Trust Framework supported the calibration of trust in algorithmic accuracy with socially responsible design principles that facilitated stakeholder participation, digital empowerment and financial well-being.

Recommendations for Financial Regulators

- Set minimum explanation criteria for high-stakes financial AI decisions.
- Require AI impact assessments and subgroup fairness testing.
- Specify documentation, traceability, and incident-reporting requirements.
- Enforce easily accessible human review, complaints, and appeals processes.
- Control third-party AI vendors and establish who is held responsible.
- Create standards for independent algorithmic auditing.
- Issue sector-specific recommendations for conditions relevant in emerging economies.

Recommendations for Financial Institutions and Auditors

- Establish multidisciplinary AI governance committees.
- Maintain complete audit trails and validation records.

- Test explanations with actual users.
- Monitor model performance and effects on vulnerable groups.
- Provide local-language and human assistance.
- Ensure reviewers possess meaningful authority.
- Analyse complaints and use feedback for system improvement.

Recommendations for AI Developers and Fintech Firms

- Integrate explainability from the design stage.
- Develop stakeholder-specific explanation interfaces.
- Disclose model limitations and uncertainty.
- Test explanation stability and actionability.
- Enable data correction and accessible audit logs.
- Design for low-literacy and low-bandwidth environments.
- Conduct continuous fairness and drift monitoring.

Recommendations for Development Agencies and Civil-Society Organisations

- Support digital and financial literacy.
- Represent marginalized groups in AI-governance processes.
- Assist users in challenging automated decisions.
- Conduct community-level impact assessments.
- Develop public-interest auditing capacity.
- Promote gender-responsive governance.
- Provide accessible information on digital financial rights.

Limitations of the Study

- The study was based on secondary evidence.
- The framework was not empirically validated.
- Emerging economies differed in institutional and technological conditions.
- Proprietary-system documentation was limited.
- Case evidence varied in quality.
- Rapid technological change could affect future applicability.

Future Research Directions

- Validate the framework through surveys and interviews.
- Test explanation formats with underserved users.
- Develop a Human-Centric XAI Trust Index.
- Conduct cross-country and gender-based comparisons.
- Undertake longitudinal studies of trust and inclusion.
- Test the framework across credit, insurance, taxation, and fraud detection.
- Examine the cost of independent AI audits.
- Develop open datasets for inclusive financial-XAI research.

References

1. Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024). *Intelligent financial system: How AI is transforming finance* (BIS Working Papers No. 1194). Bank for International Settlements. <https://www.bis.org/publ/work1194.htm>
2. Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review*, 57, Article 216. <https://doi.org/10.1007/s10462-024-10854-8>

3. Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445188>
4. Financial Action Task Force. (2021). *Opportunities and challenges of new technologies for AML/CFT*. <https://www.fatf-gafi.org/en/publications/Digitaltransformation/Opportunities-challenges-new-technologies-for-aml-cft.html>
5. High-Level Expert Group on Artificial Intelligence. (2020). *Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>
6. Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 624–635). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445923>
7. Kelly, S., & Mirpourian, M. (2021). *Algorithmic bias, financial inclusion, and gender: A primer on opening up new credit to women in emerging economies*. Women's World Banking. https://www.womensworldbanking.org/wp-content/uploads/2021/02/2021_Algorithmic_Bias_Report.pdf
8. Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: A systematic review. *Frontiers in Artificial Intelligence*, 7, Article 1456486. <https://doi.org/10.3389/frai.2024.1456486>
9. Klapper, L., Singer, D., Starita, L., & Norris, A. (2025). *The Global Findex Database 2025: Connectivity and financial inclusion in the digital economy*. World Bank. <https://doi.org/10.1596/978-1-4648-2204-9>
10. Liao, Q. V., Gruen, D. M., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI*

- Conference on Human Factors in Computing Systems* (pp. 1–15). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376590>
11. Maathuis, H., Stalenhoef, M., van Otterloo, S., Zwaal, R., van Montfort, K., & Sent, D. (2026). Designing effective explainable AI: A human-centered evaluation of explanation formats in financial decision-making. *Frontiers in Artificial Intelligence*, 9, Article 1668029. <https://doi.org/10.3389/frai.2026.1668029>
 12. Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), Article 39. <https://doi.org/10.3390/ijfs9030039>
 13. Moreira, C., Palatkina, A., Braca, D., Walsh, D. M., Leihn, P. J., Chen, F., & Hubig, N. C. (2025). *Explainable AI systems must be contestable: Here's how to make it happen* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2506.01662>
 14. Organisation for Economic Co-operation and Development. (2024). *AI principles*. <https://www.oecd.org/en/topics/ai-principles.html>
 15. Pérez-Cruz, F., Prenio, J., Restoy, F., & Yong, J. (2025). *Managing explanations: How regulators can address AI explainability* (FSI Occasional Paper No. 24). Bank for International Settlements. <https://www.bis.org/fsi/fsipapers24.pdf>
 16. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372873>
 17. Suresh, H., & Gutttag, J. V. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Article 17, pp. 1–9). Association for Computing Machinery. <https://doi.org/10.1145/3465416.3483305>
 18. Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>

19. United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
20. United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. <https://sdgs.un.org/2030agenda>