International Journal of Engineering Research & Management Technology
(Peer-Reviewed, Open Access, Fully Refereed International Journal) ISSN: 2348-4039
Volume 11, Issue-5, September-October- 2024 Impact Factor: 7.09
Email: editor@ijermt.org www.ijermt.org

Human-Centered AI Workflows for Family FinTech: Customer-Care Automation and Financial-Literacy Gamification

Naresh Kumar Paturi
Principal Architect, Greenlight Financial Technology, Inc.

Abstract. Family FinTech systems increasingly combine money movement, education, safety, and support in a single product. This paper presents a human-centered AI workflow model for such systems, grounded in my work on assistant-style support automation and engagement features that reinforce financial learning. The architecture is deliberately bounded. An AI layer classifies intent, retrieves approved policy and product guidance, answers regulated topics deterministically, routes higher-risk cases to people, and records feedback for continuous improvement. For financial literacy, I propose an age-aware design that prioritizes explainability, parental oversight, and measurable learning outcomes over raw engagement. I treat the system as socio-technical rather than as a model in isolation, and I adopt a recognized risk-management framing for AI to organize its controls. I define an evaluation methodology using call deflection, containment quality, time to resolution, unsafe-response rate, learning engagement, and learning retention, and I include a data-governance plan for minors. Numerical values, including a reported reduction of roughly fifteen percent in customer-care contacts, are drawn from this work and should be confirmed with business-analytics evidence before they are treated as generalizable results. The contribution is a safe, auditable pattern for applying AI in a child-facing, regulated financial product.

Index Terms. responsible AI, FinTech, customer-care automation, financial literacy, gamification, human-in-the-loop, data governance.

I. Introduction

Applying AI in family finance is not only a chatbot problem. Parents, children, and support agents each need trustworthy guidance, clear boundaries, and a safe path to a human when a situation calls for one. A well-designed system can reduce repetitive support work and create useful learning moments while keeping sensitive cases firmly in human hands. The design objective is help that knows its limits, because in this domain a confident wrong answer can cause real harm.

Two properties make the setting unusually demanding. First, it is regulated, so responses about money must be accurate, consistent, and explainable. Second, it is child-facing, so privacy, age-appropriateness, and parental oversight are not optional features but design constraints. The architecture proposed here treats the model as one component of a socio-technical system, with governance, measurement, and human oversight around it, rather than as the system itself.

The contributions of this paper are a bounded automation architecture in which regulated topics are answered deterministically and low-confidence cases are escalated, an age-aware financial-literacy design oriented to learning outcomes rather than engagement, a control set organized around a recognized AI risk-management framing, and an evaluation methodology with a data-governance plan for minors. The remainder of the paper reviews related work, states the model and requirements, presents the architecture and its responsible-AI controls, discusses implementation, specifies the evaluation, and closes with discussion, threats, future work, and conclusions.

II. Background and Related Work

The control framing draws on the NIST AI Risk Management Framework, which treats an AI capability as a socio-technical system to be governed, measured, mapped, and managed, rather than as a model evaluated in isolation [1].

Security and privacy controls for the surrounding system follow established catalogs [2], and payment-data handling is constrained by industry data-security requirements [3].

The retrieval-grounded design reflects the broad move toward grounding generated responses in approved sources so that the system answers from policy rather than improvising it. For the education component, public guidance on youth financial capability informs the choice of learning objectives and the emphasis on age-appropriate content and parental involvement [4]. The human-in-the-loop posture, in which automation handles routine cases and people handle ambiguous or high-risk ones, is well established in safety-critical automation.

The contribution specific to this setting is the insistence on bounded automation in a regulated, child-facing context. Regulated topics receive deterministic, policy-grounded answers; the AI is permitted to assist with routine work and learning, but it is not permitted to invent policy or to act as a financial adviser. The controls are properties of the workflow, not settings on the model, which is what makes them auditable.

III. System Model and Requirements

We model three classes of user, the parent, the child, and the support agent, interacting with a system that can answer questions, take limited actions, and deliver learning content. The AI layer has access to an approved knowledge base and to a policy engine, and it can escalate to a human queue. We assume that some requests touch regulated topics, that some involve minors' data, and that the system must produce an auditable record of what it did.

The functional requirements are to classify intent, answer from approved sources, deliver age-appropriate learning content, and escalate when confidence or risk warrants it. The non-functional requirements are that regulated topics are answered deterministically, that unsafe responses are driven toward zero, that minors' data is minimized and protected, that parents retain visibility, and that every interaction is logged for review. Table I lists the workflow components and the human guardrail attached to each.

Table I. AI workflow components and human guardrails

Component	Function	Human guardrail
Intent classifier	Identify the support or learning need	Escalate on low confidence
Approved-knowledge retriever	Ground responses in policy and product docs	No free-form policy invention
Policy engine	Answer regulated topics deterministically	Fixed responses for regulated cases
Learning-game engine	Generate age-aware questions and feedback	Curated curriculum objectives
Feedback loop	Improve routing and content	Human review of failures and edge cases

IV. Architecture and Responsible-AI Controls

The workflow has four layers: intent detection, approved-knowledge retrieval, policy enforcement, and human escalation. Financial-literacy modules use AI only within curated objectives such as budgeting, saving, trade-off decisions, and age-appropriate quizzes. The system is designed to be helpful without pretending to be an adviser, and to hand off rather than guess when confidence is low. Figure 1 shows the bounded path.



Bounded automation: regulated topics answered deterministically, low-confidence cases routed to people

Fig. 1. Bounded AI workflow. Regulated topics are answered deterministically from approved sources, and low-confidence or high-risk cases are routed to people.

A. Bounded automation

The boundary between what the AI may answer and what it must escalate is explicit. Routine, low-risk questions are handled end to end; regulated topics are answered from fixed, policy-grounded responses; and anything below a confidence threshold or above a risk threshold is routed to a human with full context. This boundary is the central safety property of the design.

B. Responsible-AI controls

Controls include age-aware content filtering, parental visibility, deterministic responses for regulated topics, prompt and response logging, bias monitoring, and explicit refusal paths for unsafe advice [1], [2]. Treating the system as socio-technical, rather than as a model alone, is what makes these controls coherent: each is a property of the workflow that can be tested and audited independently of the model's internals.

C. Learning design

The financial-literacy component optimizes for learning outcomes, not engagement. Content is age-appropriate, explanations are required rather than optional, and parental oversight is built in. The measure of success is whether a child retains a concept, not how long they stay in the app, which deliberately avoids the engagement-maximizing incentives that would be inappropriate for a child-facing financial product.

V. Implementation Considerations

The retrieval layer indexes only approved content, and the policy engine holds the deterministic responses for regulated topics so that those answers never depend on generation. All interactions are logged with the context needed for review, and data involving minors is minimized at collection and protected in storage and transit, consistent with applicable data-security requirements [3]. Escalation hands a human the full conversation and the system's reasoning so that the person starts informed rather than cold.

Quality assurance runs a held-out review set continuously to track unsafe-response rate and containment quality, and failures feed back into both routing and content. The system is built so that a regression in safety is visible quickly rather than discovered through a customer complaint.

VI. Evaluation Methodology

A rigorous study compares pre- and post-automation periods and runs an A/B learning intervention. Operational metrics include call deflection, containment quality, time to resolution, unsafe-response rate, and recontact rate. Learning metrics include engagement and, more importantly, retention measured through follow-up assessment. The evaluation must include a data-governance appendix describing how data involving minors is handled and minimized, because the validity of the result and the ethics of obtaining it are inseparable here.

The figures here are reported and illustrative values from this work. Figure 2 shows the indexed operational view, and Table II lists the metrics with their desired direction and evidence source. The reported reduction of roughly fifteen percent in customer-care contacts requires care analytics over comparable periods before it is treated as a confirmed result.

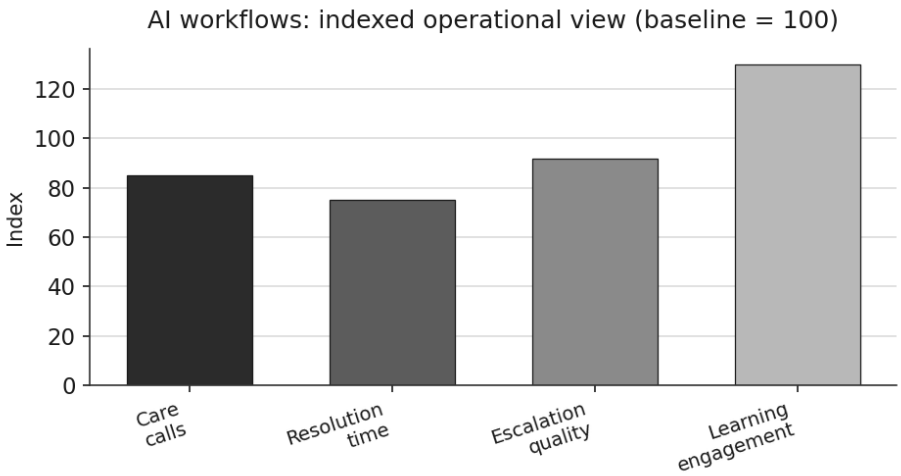


Fig. 2. Indexed operational view for the AI workflows.

Table II. Validation metrics

Metric	Desired direction	Evidence source
Customer-care contacts	Decrease (reported ~15%)	Care analytics, comparable periods
Unsafe-response rate	Near zero	Held-out QA review set
Containment quality	Increase	Disposition and recontact data
Learning engagement	Increase	Event analytics
Learning retention	Increase	Follow-up assessment

VII. Discussion and Limitations

The best AI workflow in this domain is deliberately modest. It assists with routine work and learning moments, and it slows down, explains, or escalates when a situation calls for a person. Modesty here is not a limitation; it is the safety property, and it is what allows the system to be deployed in a child-facing financial product at all.

The limitations include the difficulty of measuring learning retention rigorously without intruding on children's privacy, the maintenance burden of keeping the approved knowledge base current, and the risk that a poorly chosen confidence threshold either over-escalates, eroding the efficiency gains, or under-escalates, allowing unsafe answers through. These thresholds must be tuned against the unsafe-response metric, not against deflection alone.

VIII. Threats to Validity

Internal validity requires that deflection and engagement figures be backed by care analytics and a QA review set before they are treated as results. Construct validity requires that containment and retention be defined carefully, since deflection without quality is not a success. External validity is bounded to regulated, child-facing support and education. On confidentiality and ethics, data involving minors is minimized and protected, and any examples are sanitized. I will complete the data-governance appendix and validate these figures against care analytics before treating any number as a final result.

IX. Future Work

Future work includes a privacy-preserving methodology for measuring learning retention in minors, an analysis of confidence-threshold tuning that jointly optimizes deflection and unsafe-response rate, and a study of how parental-oversight design affects both trust and learning outcomes.

X. Conclusion

This paper presented a bounded, human-centered AI workflow for family FinTech in which regulated topics are answered deterministically, low-confidence cases are escalated, and financial-literacy content is designed for learning outcomes rather than engagement. The contribution is a safe, auditable pattern with controls organized as workflow properties and an evaluation that treats data governance for minors as part of the result. The next step is to attach analytics evidence and the governance appendix.

References

[1] NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," 2023. https://www.nist.gov/itl/ai-risk-management-framework

[2] NIST, "Security and Privacy Controls for Information Systems and Organizations," SP 800-53 Rev. 5, 2020. https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

[3] PCI Security Standards Council, "PCI DSS v4.0.1," 2024. https://www.pcisecuritystandards.org/

[4] Consumer Financial Protection Bureau, "Youth financial education," 2024. https://www.consumerfinance.gov/consumer-tools/educator-tools/youth-financial-education/

[5] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," NIST SP 800-207, 2020. https://csrc.nist.gov/pubs/sp/800/207/final

[6] M. Kleppmann, Designing Data-Intensive Applications. O'Reilly, 2017.

[7] M. T. Nygard, Release It!, 2nd ed. Pragmatic Bookshelf, 2018.

[8] Greenlight, "Debit Card for Kids and Teens," 2024. https://greenlight.com/