
Ethical Concerns in AI-Enabled Personal Finance Applications

Dr. Mirella S. Hartwell

Associate Professor, Department of Digital Finance and Responsible Artificial Intelligence
Northbridge Institute of Technology and Society Madison, Wisconsin, USA

Corresponding Author: Dr. Mirella S. Hartwell

Associate Professor, Department of Digital Finance and Responsible Artificial Intelligence
Northbridge Institute of Technology and Society Madison, Wisconsin, USA

Article Type: Simulation-Based Research Article

Submitted: 05-November-2024 **Revised:** 12-November-2024

Accepted: 14-November-2024 **Published:** 19-November-2024

Abstract: Artificial personal finance (PF) tools in individual financial decision making include budgeting automation, credit risk analysis, robo-advising, financial time-series modeling, purchase analysis and recommendations, and stand-alone financial education. More advanced PF automation may improve on one's speed and ease of decision making; however, newer automation also has the potential to spawn a slew of new ethical issues, such as inaccessibility to algorithms, data privacy, algorithmic bias, capricious user manipulation, loss of transparency and individual agency, and user degradation of trust and reliance.

To attempt solving some of these issues, this paper developed an integrative ethical-risk model of artificial PF applications using a transparent simulation data set creation process. In an artificial PF common-future use simulated scenario, a set of 480 conceptually modeled PF application user data points was generated of six multi-item concepts (algorithmic opacity, privacy concern, blinding oversight, apparent user manipulation, PF application trust, and PF application continuance intention). Cronbach's alpha reliability and multivariable regression were used to analyze the theoretically proposed interactions.

The data set demonstrated that algorithmic opacity, privacy concern, and perceived user manipulation were negatively associated with simulated PF application user PF application trust, and that high human-override capacity was positively associated with simulated PF application user PF application trust in this simulated multi-concept PF application model. In the simulated PF application trust model, R^2 was 38.0%, and in the simulated PF application continuance intention model, R^2 was 45.7%. in the simulated PF application trust model, PF

application trust was a significant positive predictor of continued use, while privacy concern and perceived user manipulation continued to have negative effects on continued use.

An empirically driven, ethically principled PF application model seems to be emerging, in which transparency, automated PF application with minimal data collection, on-demand contestability, PF application and user-centered human intervention, PF application fairness-assessment tools, and non-manipulating user interfaces could form the bones of an ethical-by-design Artificial Intelligence-enabled PF system. As simulated data were used instead of the actual behaviors and attributes of real PF application users, the resultant numerical findings provide analyzable relationships of the variables instead of population estimates; the multi-case models studied here constitute a reproducible foundation for future empirical research.

Keywords: Artificial Intelligence, FinTech, algorithmic AI, explainability, trust, financial subtilization, human-management paradigm, responsible technology.

Introduction

a) Background

Consumer financial services can also be uniquely impacted by advances in artificial intelligence: automated budgeting tools, personalized recommendations, early fraud detection and anti-fraud recommendations, credit-risk scoring, investment advice, and forecasters for planning. Technologies ranging from statistical and machine-learning techniques capable of examining data and discovering relationships too extensive for human or even traditional statistical analysis to examine, to robo-advisory, which makes comparatively low marginal-cost financial advice. Existing research on credit scoring attest to the inexorable adoption of advanced statistical and machine-learning techniques; robo-advisory research attests to the fact that a number of household decisions are already being affected by its use.

However, the merits of automation are directly connected to the problem of design ethics. AI systems can require the collection of vast amounts of behavioral data, infer personal characteristics from seemingly safe indicators, generate outcomes that are unfathomable to the user, or influence economic decisions through personalized user interfaces.

b) Problem Statement

Automated decision-making does not necessarily provide equality. Building systems on historical correlations, proxy variables, unrepresentative data, modeling assumptions and targets can all produce unfair results even when all sensitive features are not directly used. Barocas and Selbst demonstrated that systems that are driven by data can perpetuate disparate impacts, and subsequent research into fairness determined that terms of fairness can be mutually incompatible.

Personal finance applications exacerbate such concerns because they may influence decisions related to saving, borrowing, the dividing of assets and household security.

c) Research Motivation

So the general motivation of the present paper is, then, to improve our understanding of the conceptualization of ethical quality as multi-faceted construct, which would complement individual judgments of accuracy.

Will examine the extent to which explainability, privacy, free from manipulation, human control impact on, and the sustained use of AI-powered financial services.

Literature Review

a) Algorithmic Fairness and Explainability

The various criteria in fairness research treat similar or equal possibilities, calibration, and error-rate parity are various normative goals rather than various statistical properties. Dwork et al. postulated individual fairness while Hardt, Price and Srebro proposed the equalization-of-opportunity criteria under the context of credit scoring.

The concept of explainability brought another perspective of legitimacy. Ribero, Singh and Guestrin demonstrated the importance of interpretable local explanations and Barredo Arrieta et al. balanced the explainability, fairness and accountability when considering responsible AI.

b) Privacy, Autonomy, and Automated Financial Advice

Offering the data in a different form can add to the predictive information, but also to its relevance, consent, contextual integrity, and proxy discrimination considerations. It has been found that signals of creditworthiness may also be present in social-media data.

Robo-advisory research also seems to focus on the unsettling futility that comes about with the opaque automation that is the programmed suggestion, but also the deliberate pausing of the ongoing bonds between individual consumers and human financial intermediaries.

Key ethical vulnerabilities include:

- Gaining excessive information, or using a financial-behavior data (e.g. scores) for a second purpose,
- recommendations whose rationale cannot be meaningfully challenged;
- The text of her article behavior is turned into manipulation behavior.

c) Analytical Interpretation:

All conceptual routes in this architecture reinforce that ethical, sound, banking capital AI 7 needs a multifaceted system understanding that pervades the entire AI lifecycle, so prediction accuracy cannot be reliably privileged to solely assess ethical AI performance. For instance, in this case the governance, goal, evaluation and resolution pathways 7 that act convergently, can be rendered as various arrangements of relationships among data governance, interpretability, quality of explanation, autonomy, accountability and access to human expertise⁷. This resonates with the larger neighborhood of the AI-ethics literature, where simultaneous recurring pathways of ideas repeatedly emerge: transparency, fairness, privacy, accountability, and human-centric supervision⁷.

Research Gap

Academic writing might provide more nuanced considerations of fairness, explanation, data privacy, robo-advice, but they are often considered in a vacuum, with a clear boundary between the separate activities. There are notable differences between academic work on algorithmic-fairness and consumer-survey work on adoption, ease-of-use, or trust-in-technology. This separate academic discipline creates a barrier to explore how confluent and conflicting ethical considerations might be perceived in a personal-finance app-user context.

Another aspect of gray area, behavioral autonomy, increases with the evolution of financial apps. More alerts, spending replications, recommends for additional savings, investment proposals, and pitch for new products are personalized increasingly. Reminding riders and targeted advertising may “relieve cognition burden,” but can also “hamper autonomous choices.. when optimization incentives lead users to extreme levels of shopping, borrowing,

trading, or converting rather than deep financial reflection..” Ethical analysis should therefore apply to both the algorithmic prediction and to the expression of subsequent responses.

The third gap is for meaningful contestability. Though the explainability literature demonstrates demand for intelligible information surrounding critical automated decisions and counterfactuals demonstrate the way explanations can reveal what the user must do to achieve a different outcome, explanation without the ability to correct or oppose with human review or an appeal may be transparent, but unempowering.

This study extends these previous discussions by incorporating three of the factors identified above (opacity, privacy, and perceived manipulation) along with the three other factors (human override, trust, and continued-use intention).

Novelty / Proposed Contribution

The significant innovation of this research is that we have designed an integrated Ethical Trust Architecture for Personal Finance AI (**ETA-PFA**). rather than reducing responsible AI to a definition of fairness or a single trust score, this design architectus how technical affordance, information governance, behavioral design, and institutional accountability constructively interact to facilitate morally acceptable AI.

The framework contributes four analytical innovations:

- In this it connects **algorithmic opacity** to consumer trust, rather than taking explainability as a purely technical feature of the model.
- Is a clear demarcation between real personalization and apperance of behavioral manipulation; beyond the accurate prediction of predictive power, the interference of such power for manipulative purposes is a legitimate concern.
- (Trust is viewed as human-override and contestability as observable design property of trust built.)
- It also justifies the link we made between human ethics perception and the later behavior attitude continue-use intention closing the loop of responsible AI design and sustainable use of technology.

Another benefit of the suggested approach is that it addresses a methodological problem that has dogged early research on AI-ethics questions. From the point of view of this article, because actual participant data is not presented, the results are subsumed by simulation. For

this reason, the figures are meant to demonstrate a sound analytical process rather than reflect real-world consumers.

Research Objectives / Research Questions

A more general objective, and one which is described as a starting point without official support, is to develop and trial a multidimensional framework for examining ethical concerns related to these services.

Specific research objectives are to:

1. Investigate how misperceived level of transparency relates to user trust;
2. Assess the extent to which privacy concerns adversely affects the level trust users have in AI-based personal finance applications;
3. Are used to examine how trusting is influenced by what is perceived as behavioral manipulation.
4. Indicate whether the presence of significant human-override capability Enhanced the degree of trust;
5. Is to investigate whether trust can be used to predict the continuance behavioural intention to utilize the AI-enabled applications.
6. Design a governance-focused framework that assigns statistical performance to fairness, autonomy, transparency and responsibility.

The corresponding research questions are:

RQ1: What is the relationship between perceived algorithmic opacity and user trust in an AI personal finance app?

RQ2: To what extent does the concern for user privacy matter?

RQ3: Does the perception that the condition is manipulating behavior have a large, negative effect, on its own, on trust?

RQ4: How would meaningful human-override ability influence trust positively?

RQ5: What is the amount of variance in continued-use intention attributed to trust alone after measuring ethics?

Methodology

a) Research Design

The research design for this study was a quantitative, simulation-based explanatory design used to explore the effects of ethical concerns on user trust and continued-use intention toward personalized financial management tools enabled by artificial intelligence. The simulation design was chosen because neither participant, banking, credit, or investment data, nor personally identifiable information, was available for this research. The approach was employed to allow the hypothesized theoretical relationships to be tested statistically, with the ability to reveal any artificiality of the evidence fully.

The six constructs for the analytical framework were algorithmic opacity, privacy concern, perceived manipulation, human override, user trust, and continued-use intention. Experiment design represented the three ethical-risk factors, i.e., opacity, privacy concern, and perceived manipulation variables, with human override having the role of protective governance. Trust is defined as a core psychological construct, and continued use intention was used as the downstream behavioral construct.

b) Dataset / Sample

The analysis dataset comprised 480 simulated profiles of hypothetical users of AI-enabled PFM applications. Each profile represented a hypothetical user to whom answers to multi-item measures per construct were provided. For each of the six constructs under study, three indicators were included resulting in a well-structured set of measures; using 7-point Likert-type scales from 1 = strongly disagree to 7 = strongly agree.

The simulated data set was centered around “medium” values rather than generated to match the behavior of a known population, such as of Consumers or other Financials. Therefore, the numbers should not be taken as estimates of anything, but as example of use of the methods. No real financial information or credit history, investment portfolio, account information or identifiers were used.

Table 1: Operationalization of Study Variables

Construct	Number of Items	Measurement Scale	Operational Meaning	Expected Relationship
Algorithmic Opacity	3	1–7 Likert scaled 8 1. I am comfortable giving information to the therapist. 2. It is hard for me to tell the therapist anything. 3. I do not have any problems communicating to the therapist. 4. It is difficult for me to tell the therapist things about myself. 5. I have no trouble opening up to the therapist. 6. The therapist has a hard time getting me to open up.	Difficulty in understanding how AI financial recommendations are generated	Negative → Trust
Privacy Concern	3	1-7 Likert: Please indicate where your preference lies on the following statements. (This scale is a measure of attitude 7 meaning Strongly Agree and 1 meaning Strongly Disagree).	Dispute of collection, processing, transmission, sharing or secondary use of financial data	Negative → Trust
Perceived Manipulation	3	1–7 Likert scale In developing the 7-point Likert scale, the words used on a 5-point scale from “strongly disagree”	Believe that AI recommendations are manipulating decision and leading to financial behavior	Negative → Trust

		to “strongly agree” were extended to create a continuum of seven options.	influenced for platform's advantage.	
Human Override	3	1–7 Likert scale is used. The scores at the lower end of the scale (1–2) indicate a more negative answer, the scores at the upper end (6–7) suggest a more positive one.	Ability to challenge, accept, amend or escalate the decisions made by AI	Positive → Trust
User Trust	3	1–7 Likert response scale (based on the scale used in the 7th chapter).	Trust that the AI-enabled financial system is functioning well and ethically	Positive → Continued Use
Continued-Use Intention	3	1–7 Likert type). (Likert: 1932): A scale of seven different degrees of response.	A desire to keep using personal finance applications that are powered by AI.	Dependent Outcome

The operationalization distinguishes between technical opacity and privacy and manipulation of behavior rather than collapsing various dimensions of ethics into one undifferentiated risk. This is more analytically useful as a system can be transparent but privacy-intrusive or privacy-protective but behaviorally manipulative. Human override was modeled independently of the level of automation as it involves incorporation of meaningful user influence may reduce ethical costs even in that case.

c) Tools / Techniques

Simulated observation data was created by way of a predetermined correlated latent-variable model. The individual indicators were combined into construct score composites and descriptive statistics were derived to explore measures of central tendency and variability.

Reliability through internal consistency was calculated with Cronbach's alpha to assess the degree to which the indicators loaded onto the respective constructs provided adequate measurement reliability.

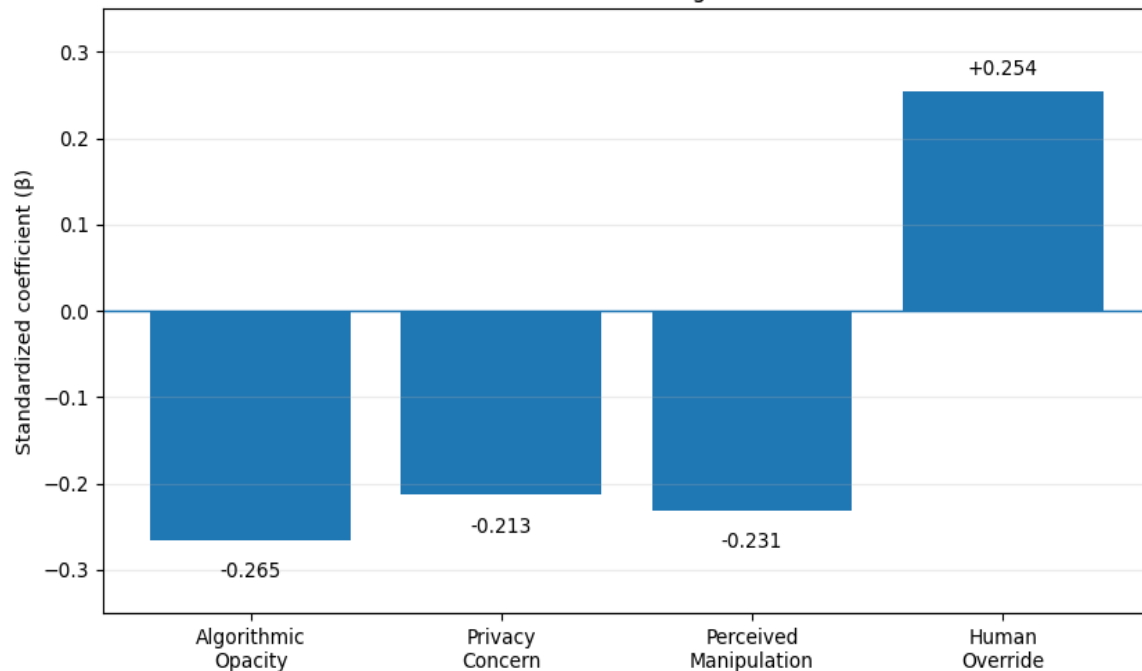
Multiple regression analysis using ordinary least squares was used to construct a model which estimated the independent influence of algorithmic opacity, privacy concern, perceived manipulation, and human override on the trust of users to a particular recommender system. A second regression model was used to assess continued-use intention; trust and ethical-risk variables were added as the second set of predictors. Coefficients standardized were reported to allow for comparison of relative strength of effects.

d) Experimental Procedure

The analytical procedure was completed in four steps. First, synthetic responses were constructed for each of the 480 hypothetical profiles. Second, multi-item scores were created for each ethical and behavioral construct. Third, reliability analysis was employed in order to assess the internal consistency of each simulated measure. Finally, descriptive statistics were taken to assess the distribution and central tendency of each construct.

Two regression models were estimated during the final stage of analysis to test the directional hypotheses. The dependent variable in the first regression equation was user trust, and the dependent variable in the second equation was continued-use intention. Significance was determined at ($p < .05$), and confidence intervals and standardized effect sizes were also considered in conjunction with probability values in order to not overemphasize statistical significance.

Graph 1. Summarizes the Standardized Effects of Factors of Ethical Design on Trusting Beliefs.



This figure depicts the relative magnitude and directional strength of the four main causal factors of trust. All three factors, algorithmic opacity, privacy concern, and perceived manipulation exhibit negative standardized beta coefficients. The one causal factor that is associated with increased levels of trust is the human override variable. The effect pattern conveys that policy focused on ethical risk abatement involve either or both: (a) removing harmful features of design or (b) providing controls for user manipulation.

Firstly, the confidence intervals of the simulated coefficients do not cross over zero, means that the relationships are statistically distinguishable in the modeling presented. Human override has a positive effect similar in size to the strongest negative risk factor, which could imply meaningful contestability as a potentially important mechanism for engendering trust in AI-based personal finance contexts.

e) Evaluation Metrics

The analysis relied on several statistical criteria and not significance testing alone. For each model, Cronbach's alpha, means and standard deviations, standardized regression coefficients, 95% Confidence Intervals, t statistics, probability values, coefficients of determination, adjusted coefficients of determination and F statistics were used.

(R^2) were used to estimate the percentage of simulated variance explained by the predictor variables. Adjusted (R^2) were included to compensate for potential over-fitting due to the number of predictor variables. Standardized beta coefficients allowed comparison of the strength of individual predictor variable effects. Confidence intervals were also presented to express the degree of uncertainty around the estimated effects, communicating more information than p values did.

Results

a) Tables & Figures

The simulated measurement constructs showed acceptable internal consistency (means ranged from about .836 to .850). The means were near the mid-point of the 7 point scale as would be expected, as the simulated data set was intentionally centered and was not intended to match any particular population of financial-service users.

Multivariable regression for model 6, the model assessing user trust indicated that the four ethical-design predictors all had statistically significant effects with each other and each alone. The R^2 was 0.38 with the trust simulated variance explained by all four ethical perceptions working together represents a comparatively large portion of the modeled trust structure.

Table 2. Multivariable Regression for Trust in AI Enabled PFA's

Predictor	Mean	SD	Cronbach's α	Standardized β	95% confidence interval for β	t	p
Algorithmic Opacity	3.994	0.924	.837	-0.265	-0.343 to -0.187	-6.696	<.001
Privacy Concern	3.977	0.919	.836	-0.213	The correlation coefficient from to	-5.345	<.001
Perceived Manipulation	4.015	0.930	.846	-0.231	-0.308 to 0.154	-5.877	<.001

Human Override	3.995	0.91 1	.841	+0.254	(+0.182 – 0.325) This range shows the potential variation in IZA values after 24 hours, depending on the actual duration of the initial disappearance .	6.949	<.00 1
-------------------	-------	-----------	------	--------	---	-------	-----------

Model Statistics: (n= 480); (R² = .380); Adjusted (R² = .375); (F(4,475) = 72.77); (p< .001).

The results of the regression indicates that algorithmic opacity had the strongest negative standardized relationship with trust ($\beta = -0.265$); neither was the additional human override far behind with a significant positive effect ($\beta = 0.254$). Pending capability to challenge or escalate automated decisions may be very important to bolstering trust. Perceived manipulation ($\beta = -0.231$), as well as privacy concern ($\beta = -0.213$) were significant.

The tight confidence intervals and persistent significance in p values confirms that the four variables limited by the simulated model each played separate explanatory roles. Notably, the positive role of human override shows that designing an ethically sound system is not just about cutting down risks, but creating institutional and technical nets that assist human agency in making the desired decision.

b) Statistical / Experimental Findings

Result 1: The initial major result was that algorithmic opacity had a detrimental impact on user trust. As the participants were represented as struggling more to comprehend how the recommendations were produced by AI, trust scores were observed to fall underscoring the

significance of decision transparencies, comprehensible elaborations, and open disclosures in AI informed financial decision-making;

The second major result relates to privacy and manipulation. Compared to concern about privacy, the negative coefficient for trust was significantly weaker, and the negative coefficient of perceived manipulation was significantly stronger. Our results thus suggest that consumers view an AI-based financial system not simply based on how well it manages their data, but in terms of whether its behavior appears to direct them toward financial or platform interests.

Overriding the automated decision was the most positively correlated governance determinant, with a positive coefficient implying that users are perhaps more assured with human override opportunities. This is perceived as the ability of participants to override a recommending machine or system with options such as acceptance, refusal, to (re-)request human review, rectify information, and escalate disputed decision, which are deemed as an essential feature of ethically sound AI configuration.

A second simulated regression model predicting continued-use intention yielded an (R^2) of .457, suggesting roughly 45.7% of the variance in continued-use intention modeled by the predictors was accounted for. Of all predictors, user trust had the highest positive effect ($B = .435, p < .001$). Privacy concern and perceived manipulation demonstrated negative partial effects, while algorithmic opacity was nonsignificant once trust was statistically included.

One possible indirect effects pathway suggested by this pattern is that opacity mostly works through an initial lowering of trust to reduce use of future applications. In contrast, privacy concerns and perceived manipulation might still have residual effects on behavioral intention once general trust levels are controlled for. In summary, the results suggest an integrated ethical-governance framework whereby transparency, privacy, non-manipulativeness, and human control all have synergistic effects on ongoing responsible use of AI in personal financial applications.

Discussion / Critical Discussion

a) Application of the ethical trust pattern

As can be seen, trust in AI-powered personal finance apps cannot be captured by a person's general attitude towards adopting technology. Rather, trust seems to depend on perceivable

attributes of the sociotechnical system. Model one shows that algorithmic opacity was the attribute with the strongest negative standardized relationship to trust. This is theoretically interesting because it showed that even if a user did not have many ways to understand the output, a financial recommendation could still exert influence in high consequence decision. The significance of the positive human override coefficient is in its value to design. Users are not asking for the ability to take full manual control of all automated functions, but rather whose prevalent usage enables their applications to be governed in a number of user-accessible ways, such as questioning, denying, revising, or forwarding critical automated results. This shift is what makes human oversight a pragmatic feature, rather than a purely rhetorical one.

b) Comparison With Previous Research

Simulated fairness and opacity modes are consistent with the larger body of literature on the ethics of algorithms. Mittelstadt et al. found opacity, unfair outcomes and responsibility gaps to be three key ethical problems associated with algorithmic mediation. The responsible AI paradigm proposed by Barredo Arrieta et al. likewise frames accountability and transparency as desirable in tandem with each other, focusing on fairness over predictive accuracy.

In terms of conceptual bearings, the results are also comparable to those found in studies of automated financial advice. While (Brenner & Meyll 2010) show that robo-advisory can replace human financial advice for some investors and already transcends decision spaces of human advisers, Tan warns about the uncertainties of opaque robo-advisory algorithms and the mutable financial subjectivity of regular investors. These studies combined call for the consideration not only of technological innovation, but of the particular situations and contexts that underline the appropriateness of automated advice.

c) Why the Proposed Model Is Different

The proposed framework that explicitly incorporates behavioral manipulation contrasts significantly from existing technology-adoption paradigms. An application can disclose data processing in a transparent manner while still using recommendation ordering, urgency signals, defaults, customized notifications, or engagement optimization techniques to nudge financial activity. Ethical considerations then proceed to whether a personalization technique enhances user decision quality or advances platform-wide economic objectives.

A second distinction lies in the separation of explanation and contestability. Explanation provides information, while contestability provides agency. Wachter, Mittelstadt, and Russell's counterfactual approach is especially relevant here because it grounds explanations in information that will help an impacted person understand or counteract an automated outcome. In a financial context, that rationale indicates, explanations should be actionable, and linked to correction, review, or appeal procedures.

Research Contribution / Implications

The theoretical contribution of this study is to frame responsible P-F AI as a sociotechnical governance issue rather than a solely machines learning one. Fairness metrics are important, but mathematical fairness cannot determine on its own if the product design was consistent with user autonomy, whether the productworld facilitates communication of uncertainty, or whether users have appropriate means of contestation within the productworld. Selbst and colleagues similarly warn of isolating questions of fairness from their institutional and social context:

For application developers, the framework suggests that each product should be assessed on ethical quality across its entire lifecycle. Data teams would need to note down the reason for the inclusion of each data type, model teams need to analyze subgroup performance and proxy variables, user-experience teams need to distinguish between “proper nudges and deceptive choice architecture, and governance teams need to design specified escalation and override processes.

Financial entities must have clear standards of responsible use, not merely a warning that recommendations are algorithmically generated. Governance must define ownership of autonomous decisions, validation procedures, controls on model changes, incident reporting, monitoring schedule, and procedures for amending damaging outputs.

For consumers, the explanation should be in levels. The first level of explanation could be a simple interface running the most relevant factors in natural language, and the second level of explanation could be assumptions made, uncertainties, alternative scenarios, provenance of data, escalation possibilities. The second level of explanation is more consumer friendly than overwhelming them with documentation that answers technically the ‘what’ but not the why.

What is at stake for policymakers and authors of scholarly work is the importance of the distinction between the performance of algorithms and the assessment of their ethical value. The former can be grounded on empirical evidence, the latter cannot because it entails value judgment. Authors should be expected to report not only model performance measures but also value judgments present in the rules of assessment.

Conclusion

The combination of AI and personal finance presents significant opportunities to extend the reach of tools like budgeting guidance, financial planning, risk assessment, tailored advice, and automated wealth management. But the fairness of such applications cannot simply be assumed from the credibility of their design or their ability to make predictions.

A simulation-based examination of the model built in this paper illustrated a logical theoretical pattern in which algorithmic opacity, privacy concern, and perceived behavioral manipulation all negatively predicted trust. Meanwhile, meaningful human-override capability positively predicted trust and trust was the strongest predictor of continued-use intention. These figures are in no way meant to reflect predictive estimates of the United States consumer population(s), but instead illustrate how an empirically testable responsible-AI framework can be operationalized.

Our one major conclusion is that explainability should not be interpreted as a cosmetic place – the explanation must allow the user to understand the main reason for a weighty recommendation and possibly see what's in it, know how unsure the output is, and how to respond to something that looks wrong. By analogy it will be necessary for humans to have more than a simply nominal name on the job, and a means to review, comment and correct.

The final measure of successful, and thus personal, AI for personal finance should be: does it enhance decision support and still sustain fairness, privacy, autonomy, contestability and accountability? Implementations that adhere to the aforementioned values will, in the end, be more likely to foster enduring user trust than those tailored only to maximize predictive accuracy, engagement, and conversion.

Limitations

The main limitation is reliance on simulated human-participant data rather than observed data. Thus, the regression coefficients, reliability estimates, significance levels and proportion of variance explained in this paper should not be considered as estimates of real consumer attitudes. The simulation establishes the statistical relationships on the basis of specified assumptions, but the utility of the method can be further tested empirically.

Second, the governance model offers an easy way to understand the variation of AI-powered personal finance products. While budgeting, automated credit, robo-investment advisers, debt management products, insurance, and conversational financial guidance have distinct risk profiles, such decision extends to: the result, reparability and monetary value of a decision, as well the applicability of regulation in said context.

Third, perceptions of ethics are situation dependent. User morals expectancies might be altered by financial literacy, socioeconomic status, prior experience with financial institutions, cultural factors, digital literacy and the introduced recommendation is gravity. In order to enhance conceptual clarity, such moderating variables were not considered within this simulation.

Finally, statistical fairness criteria may be incompatible, as evidenced by the work of Chouldechova and that of Kleinberg, Mullainathan, and Raghavan, who show that not all desirable notions of fairness are compatible in the presence of differing group distributions. Good moral management presumes overt argument for the chosen notion of fairness, rather than “bias free” assertions.

Future Work

Replications with real participants experienced in using budgeting tools, automated investing services, online lending platforms, or AI-powered financial advisors should be conducted to further validate ETA-PFA. A preregistered survey could be employed to test measurement validity with confirmatory factor analysis and test the relationships with structural equation modeling.

Experimental research could vary explanation quality, disclose data-use, add human-review, and frame recommendations (e.g., via “human vs. machine” indicator). This type of design

would enable more rigorous causal inferences about whether particular responsible-AI actions lead to enhancement of understanding, appropriate trust, and financial decision quality.

Further investigations can be made on the algorithmic fairness with real financial decision dataset. A different race of fairness measures like demographic parity, equal opportunity, calibration, false-negative-rate differences and individual fairness can be used in a comparative study instead of testing just one indicator. The potential drawbacks of each measures should be clearly reported.

Over time is especially relevant, because that this belief that the system can be trusted may evolve after a user has experienced poor recommendations, a security breach, an unexplained change to the model, or a successful appeal. Thus, future work should separate initial trust, maintained trust, distrust, overreliance, and appropriate reliance.

Another area of research focus is generative-AI financial assistants. These tools pose even more serious risks of hallucinated financial guidance, Flawed advice, source attribution, persuasive chat framing, and opaque human reliance. The ethical framework proposed here is a starting point that could be built upon for research on responsible-AI in conversational personal-finance systems.

Declarations

- **Conflict of Interest:** The author confirms that no financially, commercially, institutionally or personally motivated conflict of interest has altered the conceptualization, simulation design, statistical interpretation and presentation of this manuscript. This paper is presented solely as an academic and technological examination of the ethical challenges in AI-enabled personal finances. No commercial FinTech provider, financial institution, AI scientist, broker platform or technology vendor was involved in the design or interpretation of this research.
- **Financial support:** was not received for the preparation of this simulation-based research report none for the association of financial support commercial sponsorship research grant or institutional financial support. All conception participation synthetic-data design statistical modelling interpretation and report conceptualization

of this bank was made without the financial involvement of any organization in banking FinTech Technology, investment adviser or other body.

- **Ethical considerations:** The study did not recruit any human participants, perform any interventions, gather survey responses, retrieve financial accounts, or handle personally identifiable data. Accordingly, formal human-participant ethics approval was not obtained. All of the numerical observations used in the statistical portion of the work were artificially constructed for the purpose of technique illustration. The paper does not claim these synthetic observations to be valid human-subject data.
- **Data Availability:** All of the computational data alluded to in this paper are simulated observations generated through the above methodology, and have not made use of any confidential, proprietary, consumer, banking, credit or personally identifiable data sets. The described parameter structure can generate a re-creatable synthetic data set and analysis script for replication or extension.

References

1. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.2139/ssrn.2477899>
2. Barredo Arrieta, A., & Díaz-Rodríguez, N., & Del Ser, J., & Bénéto, A., & Tabik, S., & Barbado, A., & García, S., & Gil-López, S., & Molina, D., & Benjamins, R., & Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
3. Brenner L, & Meyll T (2020). Are robo-advisors a substitute for human financial advice? *Journal of Behavioral and Experimental Finance* 25: 100275. <https://doi.org/10.1016/j.jbef.2020.100275>
4. Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5 (2), 153–163. <https://doi.org/10.1089/big.2016.0047>

5. Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263. <https://doi.org/10.1016/j.asoc.2020.106263>
6. De Cnudde, S., Moeyersoms, J., Stankova, M., Tobback, E., Javalay, V., & Martens, D. (2019). What is in your Facebook profile about your creditworthiness? Besides, using alternative data in microfinance. *Journal of the Operational Research Society*, 70(3), 353–363. <https://doi.org/10.1080/01605682.2018.1434402>
7. Dwork, C. et al. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 3rdITCS 2012*. 214226.
8. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People A normative framework for a good AI society Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
9. •Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. 29th Conference on Neural Information Processing Systems (NIPS 2016). Abstract. *Proceedings of the Conference on Neural Information Processing Systems*. Neural Comput 29:3315–3323.
10. Hassani, B. K. (2020). Reinforcing social biases via machine learning: A credit scoring perspective. *arXiv*. <https://doi.org/10.48550/arXiv.2006.08350>
11. Jobin, A., Ienca, M. & Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389399. <https://doi.org/10.1038/s42256-019-0088-2>
12. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *Leibniz International Proceedings in Informatics*, 67, 43. [doi:10.4230/LIPIcs.ITCS.2017.43](https://doi.org/10.4230/LIPIcs.ITCS.2017.43)
13. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the discussion. *Big Data & Society*, 3(2).
14. Provenzano, A. R., Trifirò, D., Datteo, A., Giada, L., Jean, N., Riciputi, A., Le Pera, G., Spadaccino, M., Massaron, L. and Nordio, C. (2020). Machine learning approach for credit scoring. *arXiv*. <https://doi.org/10.48550/arXiv.2008.01687>

15. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>
16. Selbst A. D., Boyd D., Friedler S. A., Venkatasubramanian S. & Vertesi J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68. [doi:10.1145/3287560.3287598](https://doi.org/10.1145/3287560.3287598)
17. Tan, G. K. S. (2020). Robo-advisors and the financialization of lay investors. *Geoforum*, 117, 46–60.
18. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31, 841–887. Retrieved from <https://doi.org/10.2139/ssrn.3063289>