FairLend: Bias in AI Lending

FairLend: A Human-Centered AI Lending Dashboard to Reduce Overreliance on Biased Algorithmic Recommendations

Victor Moore

University of North Carolina at Charlotte, Charlotte, NC USA, vmoore26@charlotte.edu

Artificial intelligence is increasingly embedded in the most consequential decisions of adult life, including financial lending. Yet loan officers and other professionals tend to accept AI recommendations without meaningful critical evaluation. This overreliance is particularly dangerous in a domain with a well-documented history of racial redlining, proxy discrimination, and systemic exclusion of marginalized communities from access to capital. This paper presents FairLend, a human-centered AI lending dashboard designed to apply cognitive forcing functions and augmented sense-making to help loan officers critically evaluate algorithmic recommendations and reduce overreliance on potentially biased AI outputs. Drawing on prior work on algorithmic bias, human-AI decision-making, and cognitive intervention design, I propose and build a two-condition comparative prototype. Condition A and Condition B. The dashboard is presented as the primary research artifact. Implications for human-centered AI design in high-stakes domains are discussed.

CCS CONCEPTS • Human-centered computing → Human computer interaction → HCI design and evaluation methods • Applied computing → Law, social and behavioral sciences

Additional Keywords and Phrases: algorithmic bias, cognitive forcing functions, human-centered AI, financial lending, overreliance, augmented sense-making, proxy discrimination, redlining

ACM Reference Format:
Victor Moore. 2026. FairLend: A Human-Centered AI Lending Dashboard to Reduce Overreliance on Biased Algorithmic Recommendations. In HCAI Course Project, Spring 2026. University of North Carolina at Charlotte.

1 INTRODUCTION

Some of the most important needs we have as adults are now being shaped by the reach of AI. AI is being used in the decision-making process for things such as jobs, apartment seeking, insurance, and financial lending [1]. The advantages AI affords users are seemingly boundless. Yet we must ask: are we actively considering the potential drawbacks? This question is especially important in the realms listed above, where the stakes could not be higher.

In the domain of financial lending, AI and machine learning models are used to score applicants and generate approval recommendations. These models are trained on historical data that reflects decades of discriminatory lending practices, racial redlining, and systemic exclusion of marginalized communities from capital [5, 7, 13]. When a model learns from this data, it does not learn to be fair -- it learns to replicate the patterns it was trained on. The inputs it uses, including zip code, income, and debt-to-income ratios, are correlated with race and gender through historical inequity. The model therefore produces racially and socioeconomically biased outputs even when race and gender are never explicitly included as variables [2, 3, 12].

This problem is compounded by a well-documented human tendency to trust algorithmic recommendations over human judgment [12]. Research shows that people are less likely to perceive AI systems as equally (or more) biased than human decision-makers, precisely because algorithms appear to operate on neutral mathematical logic [3]. This creates a dangerous feedback loop: biased models produce biased recommendations, professionals trust those recommendations without questioning them, and the discrimination embedded in historical data is perpetuated and amplified [8].

This paper presents FairLend, a human-centered AI lending dashboard designed to interrupt this feedback loop. FairLend applies cognitive forcing functions, interface-level interventions that require users to engage in deliberate analytical thinking before acting on AI recommendations [4], to the specific context of loan review. The dashboard is grounded in the principle that augmented sense-making, not AI autonomy, is the appropriate model for high-stakes decision environments.

The central research question guiding this work is: How can a human-centered AI lending dashboard leverage cognitive forcing functions and augmented sense-making to help loan officers critically evaluate algorithmic recommendations and reduce overreliance on potentially biased AI outputs?

The remainder of this paper is structured as follows. Section 2 reviews prior work on algorithmic bias in lending, human overreliance on AI, and cognitive forcing functions. Section 3 describes the methodology used to design the FairLend dashboard. Section 4 presents the dashboard as the primary research artifact. Section 5 discusses the implications of this design. Section 6 addresses limitations and future work. Section 7 concludes.

2 BACKGROUND AND RELATED WORK

2.1 The History of Bias in Financial Lending

Financial lending has never been an equitable system. Access to capital has historically been contingent on social and economic markers that have consistently excluded Black Americans and other marginalized communities. Financial lending has long benefited a selective segment of society. An applicant has historically had to meet specific social and economic criteria in order to access capital: criteria that have knowingly and consistently disadvantaged marginalized people. History begets our current situation.

The practice of redlining, in which the federal government and private lenders systematically denied mortgage credit to residents of predominantly Black neighborhoods, is among the most well-documented examples of institutionalized financial discrimination in United States history [7]. Even after the Fair Housing Act of 1968, the structural disadvantages created by redlining did not disappear. Individuals from historically redlined neighborhoods continue to face higher loan denial rates, restricted access to resources, and greater exposure to predatory lending practices, despite being financially qualified [5]. The legacy of structural racism in lending is measurable not only in financial terms but in health outcomes, intergenerational wealth accumulation, and access to opportunity [13]. Systematically charged discrimination exists in our most essential arenas: education, employment, health care. It impacts how resources are distributed and who can access them [13]. Marginalized people still fight the harsh realities of being othered based on their neighborhoods, resulting in divestment and the persistent existence of poverty. Blockades to mortgage credit, a continued influx of predatory lending, and higher loan denial rates plague people from these communities [5].

The introduction of AI into this already compromised system carries the significant risk of further encoding and automating historical discrimination. The financial sector is heavily regulated and governed by policy and law, and yet biases have always existed that impact marginalized peoples [1]. Today we must ask: have we chosen technological advancement over equitable evolution in the financial realm?

2.2 Algorithmic Bias and Proxy Discrimination

Machine learning models used in lending are built on historical data, and that data carries the biases of the systems that produced it [1]. When a model is trained on loan approval records that reflect decades of discriminatory practice, it learns to associate creditworthiness with variables correlated with race and socioeconomic status. This is not because those variables are causally relevant, but because historical discrimination made them statistically predictive in a biased dataset. This phenomenon is known as proxy discrimination [2]. Zip code is a canonical example: because of housing segregation, a borrower's zip code is strongly correlated with their race. A model that uses zip code as an input effectively discriminates by race without ever encoding race as a variable.

Data training and collection can be structured in ways that allow stakeholders to avoid responsibility for the results that come from algorithm reliance. Data does not have to directly reference race to produce racially disparate outcomes. Our digital footprints are vast: people can be discriminated against based on social media activity or browsing behavior and would never know why they were denied various opportunities. As Barocas and Selbst

observe, the problem does not rest within the algorithm alone. It exists at the root: the preexisting historical inequalities and injustices that have always permeated our society [2]. Biases are increasingly perpetuated by big data. We see this in various arenas: media, medicine, surveillance, and law.

The financial sector has been identified as underserved in terms of research specifically addressing and seeking to eliminate bias in AI systems [15]. As of August 2023, this body of work remained limited relative to the scale of the problem [15]. Akinwumi et al. [1] provide a policy agenda for federal financial regulators, arguing that AI and ML models perpetuate the land mines that have always existed in this space. Customers are the victims here, susceptible to vulnerabilities that powerful financial institutions are capitalizing on through the implementation of biased algorithms. Sophistication of these models is increasing, yet attention to the people they are impacting seems to be an afterthought.

Winder, Hildebrand, and Hartmann [17] demonstrate that large language models exhibit availability bias and recency bias in investment contexts, favoring data that is more readily accessible and recent. This is a pattern that systematically disadvantages less-documented populations. Companies such as Scienaptic, Zest AI, and TurnKey Lender have developed ML models that assist lending procedures. These AI-lending tools can augment some procedural steps; yet allowing these models complete autonomy is irresponsible and ethically dangerous. Accountability must rest on humans, not LLMs [18].

A critical mechanism through which bias operates quietly in AI lending systems is the approval threshold: the numeric cutoff that determines whether a machine-generated score results in an approval or a denial. The threshold itself is a human decision, calibrated by institutions, and rarely scrutinized with the rigor applied to other lending criteria. If an applicant's score is depressed by proxy variables like zip code, and the threshold disproportionately excludes applicants who score lower for proxy-related reasons, the result is systemic discrimination encoded in a number. No policy explicitly discriminates. The threshold does the work quietly.

2.3 Human Overreliance on Algorithmic Recommendations

Biases exist, yes, even within AI systems. This is seemingly a major surprise for many machine learning enthusiasts. Big data must be scrutinized for the overreliance it affords us in an increasingly algorithm-driven world [12]. We have been shown to trust ML systems blindly without knowledge of how their answers are generated, how our data is consumed, or just how problematic these systems can be.

Research has shown that users are less likely to deem AI as biased [3]. This is due to the fact that people look at these systems as strictly quantitative computers, unaffected by the issues that exist within human life. The reality is that such blind acceptance has the potential to further perpetuate the mistreatment of marginalized groups and further limit their access to necessary funding, housing, and care. Professionals in finance are vulnerable to exacerbating the problem by allowing systematic bias to go unchecked in the lending qualification process.

The opaque nature of most AI systems makes it difficult for users to ascertain the mistakes these models make, why they are made, and how they occur [14]. Different from tailored augmentation, users may decide to delegate complete authority over decision-making to the AI model at hand. AI is often touted for its ability to remain unaffected by the things that limit humans: fatigue, inconsistency, emotion. Seemingly, less effort is required of AI systems; which, increases some people's reliance on systems they believe can do no wrong.

Moreover, research shows that people have the capacity to rely even more heavily on AI in situations where the AI is presented with data outside the scope of what it was trained on [6]. Within these scenarios, AI recommendations can be more likely to be erroneous, and yet users trust them more, not less. This is a critical vulnerability in any high-stakes AI-assisted decision environment. As of November 2025, finance, asset, and wealth management services are among the highest adopters of AI [18], making this dynamic particularly urgent in the lending context.

Research from the American Psychological Association found that algorithmic decisions yielding gender or racial disparities are less likely to be seen as discriminatory, because people believe algorithms apply rules without regard for individual characteristics [3]. This belief is factually incorrect but widespread, meaning the professionals most responsible for catching discriminatory outcomes are the ones least likely to perceive them when an algorithm is involved. It is dangerous to look at algorithms as just cold, code-based technology that only makes decisions based on mathematical logic. Even if that were true, if the input is skewed, the output will be too. These systems do not possess the ability to evade harmful decision-making in favor of equitable outcomes.

Humans are also likely to learn and adapt the biases they are exposed to in AI systems [8]. It has been shown that humans often do not realize they can be affected by the biases embedded in these systems. This lack of realization ironically makes them more vulnerable to those biases. Biased algorithms produce biased evaluations [8]. There is seemingly an arms race occurring in the AI space, with growth and development as the focus, and this creates problems for those often left out of the room.

2.4 Cognitive Forcing Functions and the Psychology of Decision-Making

Surprisingly, research shows that designs which seek to reduce overreliance on AI through explanations displayed in the UI are not as effective as one might hope. Simply adding explanations to AI recommendations may even increase overreliance [4]. Explainable AI should not be abandoned entirely. Further research may be needed to explore its full potential and the circumstances under which it proves most helpful. However, the evidence prompted a different approach in this work.

Buçanca et al. [4] propose cognitive forcing functions as a more effective intervention. A cognitive forcing function is a design element that compels the user to assess from a place of reason rather than instinct. Tools such as checklists can serve this purpose; the goal is to force the user to engage deliberately before acting [4]. Their study found that cognitive forcing interventions significantly reduced overreliance on AI recommendations compared to simple explainable AI approaches. Although, users rated these designs as less favorable, reflecting the known tension between what produces better outcomes and what users prefer in the moment [4].

Since the results of this kind of research touch on human thinking patterns, a psychological approach is necessary. We as humans rely on patterns we have created to complete daily tasks. In high-stakes realms such as finance, analytical thinking is essential. It is far easier to accept the recommendations of AI without pushback when the AI appears to check all the boxes. My work is grounded in the belief that people can be prompted to slow down and perceive with depth, to deploy System 2 level thinking, deeper analysis, critical reasoning, and deliberate decision-making, even in the face of a technological ecosystem that increasingly incentivizes quick, algorithm-driven responses [10].

This tension has precedent in medical decision-making research. Dual-process cognitive interventions have been studied as a method for reducing diagnostic error in clinical settings [11]. Sherbino et al. found that while cognitive forcing strategies showed promise in theory, their effectiveness was inconsistent in controlled trials [16], underscoring the need for carefully designed implementations that account for context and workflow. Research on AI-supported note-taking further supports this direction: greater AI automation reduces cognitive engagement, and although results were not favorable under fully automated AI conditions, participants still favored the AI that provided the highest level of ease and the lowest cognitive effort [19]. Convenience has been proven to not produce cognitive benefits. In high-stakes domains, interface design must actively resist the path of least cognitive resistance.

The principles and limits of algorithm-in-the-loop decision-making are also relevant here. Algorithms may not display reflexivity: the kind of reasoned, contextual judgment that accounts for nuance and individual circumstance [20]. Nuance is trained and learned. Research shows that it is currently not a default, out-of-the-box guarantee in AI systems [20]. In lending, this loss of reflexivity has direct consequences for applicants whose profiles do not conform

to the majority patterns embedded in training data. The lending space requires empirical data as well as experiential reasoning. Humans excel at empathy, compassion, and the ability to relate to the human experience. In a larger research study, one may ask the question: is algorithm-in-the-loop beneficial enough to justify inclusion in high-stakes arenas? There are a great deal of things that must be considered, law, politics, ethics, before adopting these systems. Are there arenas where including AI is not sensible or safe [20]?

Hassabis et al. [9] further argue for the necessity of incorporating neuroscience into AI development, suggesting that understanding human cognition is essential to designing the interfaces through which humans and AI collaborate. The motivation for their work differs from what is studied here, yet it is relevant because it gives further evidence of the importance of studying truly thoughtful, intentional augmentation.

3 METHODOLOGY

3.1 Research Approach

This project follows a research-through-design methodology. The primary contribution is a designed artifact, the FairLend dashboard, developed through an iterative process grounded in the literature reviewed in Section 2. Each design decision embedded in FairLend is traceable to specific findings in prior work. This project was carried out as a solo research and design effort. I identified the topic, conducted the literature review, designed both dashboard conditions, built interactive prototypes in Figma, and prepared all study materials independently.

3.2 Pivot from Explainable AI to Cognitive Forcing Functions

My initial idea was to explore the methods of explainable AI. I thought that providing context for the AI recommendation would serve as enough friction to cause the user to give further thought before blindly accepting the AI's suggestion. As I researched further, I found that this method had already been more thoroughly explored, and the evidence did not support my hypothesis. Research by Buçanca et al. [4] demonstrated that explanations can increase rather than decrease overreliance, because a persuasive explanation provides users with additional justification for deferring. After reviewing this data, I decided to experiment with cognitive forcing functions. The psychological reasoning behind this method seems sound and, in my assessment, more promising.

3.3 Dashboard Design

Two dashboard conditions were designed and prototyped to enable a comparative study of the effect of cognitive forcing interventions on loan officer decision-making. Both dashboards display the same fictional applicant, Marcus T. Williams, applying for a personal loan of $24,500 for home improvement, to ensure comparability across conditions. Both were built as interactive prototypes in Figma using Material Design principles and the IBM Plex Sans typeface.

Condition A, referred to as LendCore, represents a standard AI lending dashboard as it currently exists in practice. It presents the applicant's profile and financial data alongside an AI-generated risk score and a binary approve or deny recommendation, with action buttons available immediately. There are no required steps, no prompts for independent judgment, and no bias flags. This condition operationalizes the status quo described in the literature: a design that maximizes efficiency and minimizes cognitive friction at the cost of critical engagement.

Condition B, referred to as FairLend, is the human-centered dashboard proposed by this research. It is structured as a four-step process, with each step representing a cognitive forcing intervention adapted from Buçanca et al. [4]. The design of this dashboard and its four steps are described in detail in Section 4.

3.4 Study Design

The study was designed as a small-scale, qualitative comparative study involving between three and five participants. Each session was planned to last approximately 25 to 30 minutes and to consist of a welcome and introduction, a

Condition A session using the think-aloud method, completion of a post-task survey, a Condition B session with verbal responses collected at each text input, completion of a second post-task survey, and a debrief. Due to time constraints inherent in a solo semester-long research project, the planned user study was not conducted within the scope of this work. The FairLend dashboard is presented here as the primary research artifact, with the study design documented for replication in future work.

4 RESULTS

4.1 The FairLend Dashboard as Research Artifact

The primary result of this research is the design and implementation of the FairLend dashboard as a working Figma prototype. The artifact operationalizes three cognitive forcing interventions within a lending-specific interface designed to make bias visible and overreliance structurally difficult. Each element of the design is a deliberate response to a specific problem identified in the literature. In lieu of a completed user study, the FairLend dashboard is presented here as a working prototype. It was generated from the research within this paper, and I believe it is a strong solution grounded in that foundation. That said, user research would be required to push things further; user testing would be necessary to validate or invalidate the proposed mechanisms.

Figure 1 shows Condition A: the LendCore standard dashboard. The AI recommendation of deny is presented prominently alongside a score of 34 out of 100. Three action buttons, Approve, Deny, and Hold, are immediately available. A model performance panel shows an officer override rate of 3.2 percent, meaning that loan officers accept the AI recommendation 96.8 percent of the time. This single statistic captures the core problem that FairLend is designed to address. The same flag noting that zip code 30310 is present in the applicant data appears in the left sidebar, yet nothing in the interface requires the officer to engage with it. In Condition A, this information is available but meaningless. It is an opaque model that gets straight to the point. What the officer sees most clearly is the AI score and the recommendation to approve or deny. I see in this dashboard an example of the overreliance on AI that this research seeks to address.

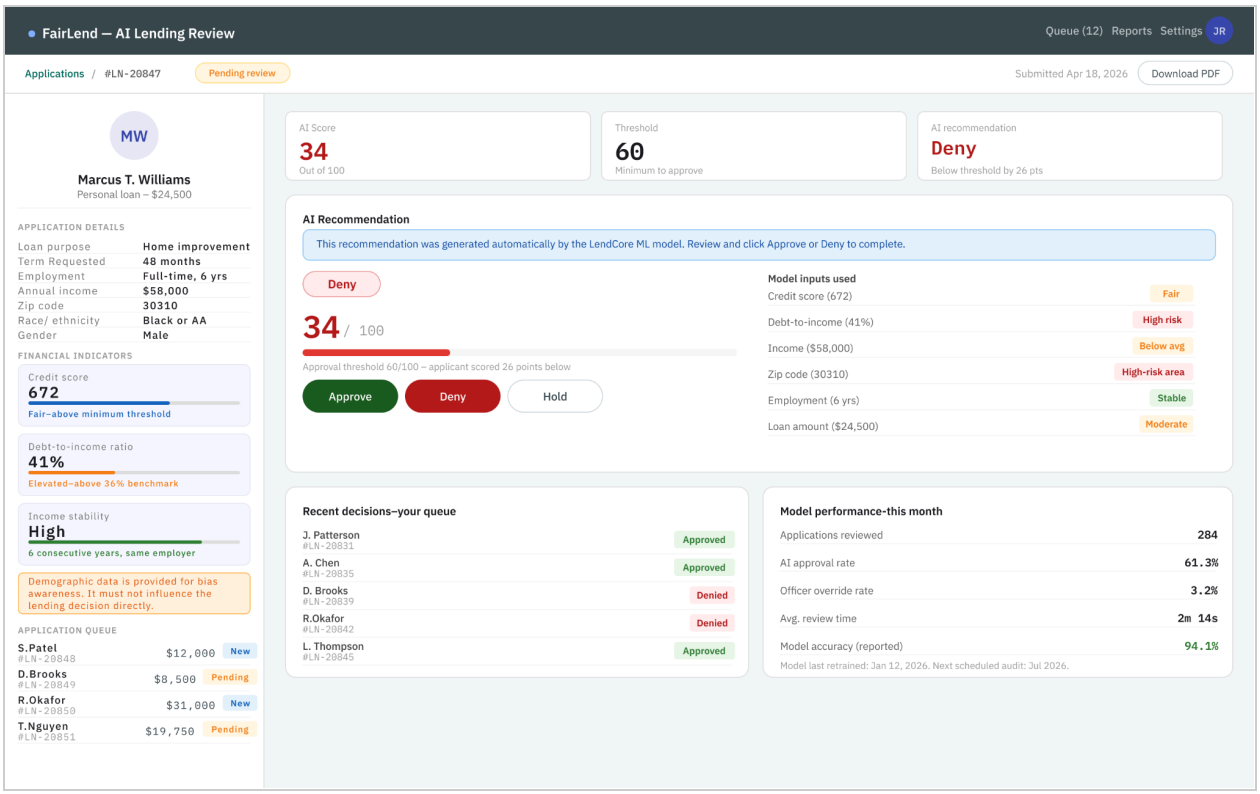

Figure 1. Condition A – LendCore Standard Dashboard. The AI recommendation and action buttons are immediately visible with no required steps before a decision can be made. The officer override rate of 3.2% reflects the degree to which loan officers follow AI recommendations without intervention.

4.2 The FairLend Four-Step Design

Figure 2 shows the FairLend dashboard across its four steps. The interface is built on cognitive friction. In order to submit a decision on an applicant, the lending professional must work through a series of prompts designed to engender deliberate cognition. The dashboard color system communicates trust (blue), caution (amber), alert (red), and completion (green): a visual language designed to reflect the weight appropriate to each phase of the process.

Step 1 – Initial Judgment. Before the AI recommendation is visible, the officer must review the applicant's financial profile and demographic data and record an independent preliminary judgment: lean toward approve, lean toward deny, or unsure. A written explanation of their reasoning is also required. The user cannot see the AI recommendation until this step is complete. This operationalizes the delayed reveal intervention [4], forcing an independent reference point before the AI's output can anchor the officer's thinking. Human-in-the-loop is presented here, responsibility and ownership over the decision-making process is reinforced at the very beginning. A flag visible in the left sidebar reads: 'Demographic data is provided for bias awareness. It must not influence the lending decision directly.' This places bias awareness at the entry point of the process, before the AI recommendation is ever introduced.

Step 2 – AI Recommendation. Upon completing Step 1, the AI recommendation is revealed as a medium risk tier with a confidence score of 54 out of 100. The score is situated within a zone system identifying deny, review, and approve thresholds, explicitly framing 54 as falling in the review zone where human judgment is required. This step presents further prompting and transparency: six contributing factors are displayed with colored tags, and zip code 30310 is flagged as a high-risk area negative contributor. A reflection panel below the recommendation presents the officer with a direct prompt: does the AI's reasoning align with your own assessment from Step 1? A written response is

required before proceeding. This operationalizes the prompted reflection intervention: the officer must articulate where their judgment agrees and disagrees with the algorithm before they can move forward.

Step 3 – Bias Review. This step presents the most direct intervention against proxy discrimination in the dashboard. Two bias flags are surfaced: the first identifies zip code 30310 as a known proxy for race in lending models; the second identifies the applicant's debt-to-income ratio as potentially reflecting a gender wage gap in the model's income baseline. The officer must select one of three acknowledgment options before proceeding: to factor the flags into the final decision, to determine that the flags do not materially affect the case, or to escalate the application for review due to concerns about bias. This represents human-centered AI in what I believe is its most important form: a system that recognizes and admits its own potential for bias, and then transfers responsibility for navigating that reality to the human in the loop. The officer is given choices on how to proceed; the goal is deeper cognition.

Step 4 – Final Decision. The officer is presented with a summary of their journey through the prior three steps. Four decision options are available: approve, deny, escalate to senior review, or approve with conditions. The application can also be flagged for compliance or saved to return to later. A required reasoning field must be completed before submission. Whatever the decision is, reasoning is required to be documented. I spoke about responsibility resting on the loan officer throughout this paper, and this step exhibits exactly that, through its requirements. The dashboard makes it impossible to intentionally or unintentionally relegate responsibility to a biased AI system. An audit trail is being built with this and every decision made. This is relayed to the user before they complete the final step.

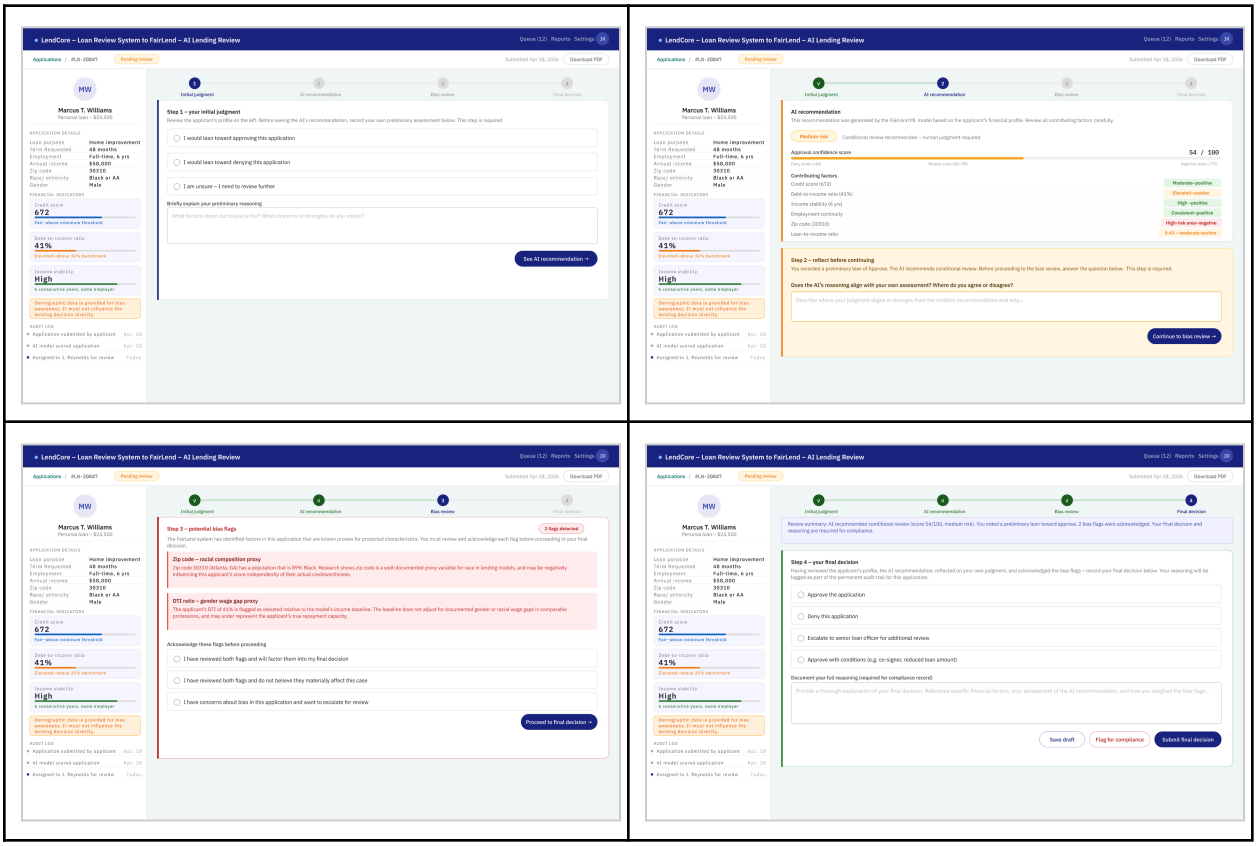

Figure 2: Condition B – FairLend Human-Centered Dashboard. Clockwise from top left: Step 1 (Initial Judgment), Step 2 (AI Recommendation), Step 3 (Bias Review), Step 4 (Final Decision). Four required steps move the loan officer through independent judgment, AI review, bias acknowledgment, and documented decision before any action can be taken.

4.3 Contrasts Between Conditions

The standard dashboard has a similar panel on the left side. The space is utilized differently: specifically at the bottom of the left panel, where the officer sees an application queue instead of the audit log present in Condition B. The most the officer gets by way of transparency in the LendCore model are the contributing model inputs. What is noteworthy is the officer override rate and the average review time. What I observe in Condition A is an interface that makes overreliance easy and critical evaluation hard. The fact that the AI recommendation offers a small degree of insight into how the score is compiled, without explaining or breaking it down, may make the officer more likely to trust it, not less. In Condition B, that same information is broken down, named, and contextualized with bias flags. The AI is repositioned from authority to input. That repositioning is the core design contribution of this work.

5 DISCUSSION

FairLend addresses the research question by operationalizing three cognitive forcing interventions within a lending-specific interface designed to make bias visible and overreliance structurally difficult. Each intervention corresponds to a specific failure mode identified in the literature. In an era heavily influenced by algorithms, I believe that true intervention is needed in order to dissuade the automatic adoption of information received from AI models. We have to save ourselves. The kneejerk acceptance of Big Data is far too advantageous to those who create and push these algorithms [12]. We cannot simply wait for institutions to course correct.

The delayed reveal in Step 1 addresses the anchoring problem: when an AI recommendation is shown first, it establishes a reference point that subsequent human judgment tends to align with rather than independently evaluate [4]. By requiring the officer to commit to a preliminary position before seeing the AI's output, the design creates an independent reference point. Even if the officer ultimately aligns with the AI's recommendation, they have engaged in the deliberate reasoning process that Buçanca et al. [4] identify as the mechanism through which cognitive forcing functions improve decision quality. Causing a user to slow down and think more deeply is an approach I believe will greatly lessen the temptation to overrely on algorithms.

The prompted reflection in Step 2 addresses the finding that explanations can increase overreliance when they are persuasive [4]. Rather than simply displaying the AI's reasoning, FairLend requires the officer to articulate their own response to it in writing. This externalization of reasoning is itself an intervention: writing requires more deliberate processing than passive reading. And the act of articulating disagreement, makes it more cognitively available as an option.

The friction gate in Step 3 addresses proxy discrimination specifically. The bias flags do not change the AI's recommendation. They change what the officer knows about what drove it. Having acknowledged that zip code 30310 is a racial proxy that may be inflating the risk score, the officer cannot claim ignorance of that dynamic in their final decision. The required acknowledgment creates accountability without removing human authority. This is what the human in the loop should mean: not a concept, but a structural requirement embedded in the interface itself.

The documented reasoning requirement in Step 4 addresses the accountability gap in current lending practice. Standard dashboards like Condition A record the final decision but not the reasoning behind it. FairLend creates a compliance record that includes the officer's preliminary lean, their reflection on the AI recommendation, their acknowledgment of the bias flags, and their full documented reasoning. This audit trail makes the human decision-making process legible in a way that the current standard does not.

This design is consistent with the call by Akinwumi et al. [1] for regulatory frameworks that ensure AI tools in lending advance financial inclusion rather than undermine it. A dashboard that requires loan officers to acknowledge bias flags and document their reasoning creates the kind of institutional record that regulators, compliance officers, and courts can examine. It does not solve the problem of biased training data. But it creates conditions under which biased outcomes are more likely to be identified, questioned, and corrected.

The study of Human-Centered Artificial Intelligence demands that we get realistic in our pursuits. Otherwise, the name of this field of study is intriguing at best, idealistic at worst. This research intends to contribute potential guardrails that leverage psychology, behavioral science, and transparency in service of the people that these systems most affect.

6 LIMITATIONS AND FUTURE WORK

This research has several important limitations that should be acknowledged openly. I am a novice in the realm of financial lending systems and am not a finance professional or an AI expert. I am a student, new to the field of Human-Centered AI. This project is a launch point for deeper investigation.

The planned user study was not completed within the scope of this work. Running this project solo made executing the experiment within the semester timeline difficult. My scope and perspective alone are limited, and the absence of a research team constrained both the breadth of the study design and my ability to carry it out. With further development, I would like to actually run an A/B testing experiment. Even if not conducted with lending professionals, the goal would be to capture participants' responses to both dashboards in real time. Running this experiment with professionals in the field and with true administrative access to existing lending dashboard systems would add significant validity.

The study participants that were planned were not lending professionals. The extent to which FairLend would produce similar effects among actual loan officers (with domain expertise, institutional incentives, and the time pressures of a real lending workflow) remains an open empirical question. This project has the potential to be far more impactful if experts in the field can be included as participants.

Additionally, I researched what I was able to access regarding existing lending dashboards. A more comprehensive competitive analysis of platforms such as Scienaptic, Zest AI, and TurnKey Lender, with access to their actual interfaces and documentation, would have strengthened the design rationale for FairLend.

FairLend addresses the human decision-making layer of the bias problem but does not address the upstream problem of biased training data. A dashboard that requires officers to acknowledge bias flags is more protective than one that does not, but it operates downstream of the model that generates the biased score in the first place. Structural solutions, including regulatory oversight of model training data, algorithmic auditing requirements, and fairness-aware machine learning methods, are necessary complements to the human-centered design approach taken here.

Future work should include a larger-scale study with lending professionals in a realistic institutional setting. Research should also examine whether the effects of cognitive forcing interventions persist over time as officers grow familiar with the dashboard, or whether familiarity breeds a new form of automation bias in which officers complete the required steps without genuine deliberation. The specific formulation of the bias flags should be refined in collaboration with experts in fair lending law and algorithmic auditing.

7 CONCLUSION

This paper has argued that the standard design of AI lending dashboards is a site of harm. This is not because the AI is malicious, but because the interface design encourages passive acceptance of algorithmic recommendations that are themselves products of biased historical data. The problem is not simply that AI is biased. The problem is that the interfaces through which professionals interact with AI make overreliance easy and critical evaluation hard. Algorithmic systems do not encompass a level of reliability that can be trusted when asked to perform beyond how they were trained. Despite the rich potential these systems promise, this reality cannot currently be escaped.

FairLend proposes a different design philosophy: in high-stakes domains, the role of the interface is not to minimize cognitive effort but to structure it. The four-step process (independent judgment, AI review, bias acknowledgment, and documented reasoning) does not remove the AI from the process. It repositions the AI as one input in a human-led reasoning process, rather than the conclusion of a process that humans merely ratify. My project presents structured augmentation. This prototype dashboard directly opposes giving AI autonomy in high-stakes scenarios such as finance.

The people most likely to be harmed by algorithmic bias in lending, Black Americans and other marginalized communities already subjected to centuries of financial exclusion, are the people who can least afford to be further disadvantaged by the tools that are supposed to make the system more efficient. The study of Human-Centered AI demands that we get realistic. *We have to save ourselves.* FairLend is one step toward taking that responsibility seriously at the level of the interface, and toward ensuring that the human in the loop is not simply a concept, but a commitment.

ACKNOWLEDGEMENTS

I thank Prakasam Venkatachalam, who graciously allowed me to take on an idea that he first presented in our Human-Centered AI class at the University of North Carolina at Charlotte, Spring 2026. Thank you to Professor Cori Faklaris for her instruction in the field of Human-Centered AI and to teaching assistant Narges Zare for her assistance with paper structuring and ACM formatting. My Spring 2026 Human-Centered AI cohort at large, was instrumental in my growing understanding of this field.

Claude AI, developed by Anthropic, was used as an ideation and design tool throughout this project. Claude assisted in the conceptualization and prototyping of the FairLend and LendCore dashboards, and served as a thinking partner during the design process. All research, writing, literature review, and intellectual decisions reflected in this paper are my own.

REFERENCES

[1] Michael Akinwumi, John Merrill, Lisa Rice, and Maureen Yap. An AI fair lending policy agenda for the federal financial regulators.

[2] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *SSRN Electron. J.* (2016). https://doi.org/10.2139/ssrn.2477899

[3] Andrea Bonezzi and Massimiliano Ostinelli. 2021. Can algorithms legitimize discrimination? *J. Exp. Psychol. Appl.* 27, 2 (2021), 447–459. https://doi.org/10.1037/xap0000294

[4] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–21. https://doi.org/10.1145/3449287

[5] Liang Chen, Bruce C. Mitchell, Sarah E. Laurent, Jad K. Edlebi, and Helen C.S. Meier. 2026. Historical redlining and contemporary home mortgage lending: Investigating neighborhood-level racial disparities in mortgage originations in the United States. *Cities* 170, (March 2026), 106632. https://doi.org/10.1016/j.cities.2025.106632

[6] Chun-Wei Chiang and Ming Yin. 2021. You'd Better Stop! Understanding Human Reliance on Machine Learning Models under Covariate Shift. In *13th ACM Web Science Conference 2021*, June 21, 2021. ACM, Virtual Event United Kingdom, 120–129. https://doi.org/10.1145/3447535.3462487

[7] Price Fishback, Jonathan Rose, Kenneth A. Snowden, and Thomas Storrs. 2024. New Evidence on Redlining by Federal Housing Programs in the 1930s. *J. Urban Econ.* 141, (May 2024), 103462. https://doi.org/10.1016/j.jue.2022.103462

[8] Moshe Glickman and Tali Sharot. 2025. How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nat. Hum. Behav.* 9, 2 (February 2025), 345–359. https://doi.org/10.1038/s41562-024-02077-2

[9] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. 2017. Neuroscience-Inspired Artificial Intelligence. *Neuron* 95, 2 (July 2017), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011

[10] Radwa Khalil and Martin Brüne. 2025. Adaptive Decision-Making “Fast” and “Slow”: A Model of Creative Thinking. *Eur. J. Neurosci.* 61, 5 (2025), e70024. https://doi.org/10.1111/ejn.70024

[11] Kathryn Ann Lambe, Gary O’Reilly, Brendan D. Kelly, and Sarah Curristan. 2016. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual. Saf.* 25, 10 (October 2016), 808–820. https://doi.org/10.1136/bmjqs-2015-004417

[12] Jennifer M. Logg, Julia A. Minson, and Don A. Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* 151, (March 2019), 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

[13] Emily E. Lynch, Lorraine Halinka Malcoe, Sarah E. Laurent, Jason Richardson, Bruce C. Mitchell, and Helen C.S. Meier. 2021. The legacy of structural racism: Associations between historic redlining, current mortgage lending, and health. *SSM - Popul. Health* 14, (June 2021), 100793. https://doi.org/10.1016/j.ssmph.2021.100793

[14] Samir Passi and Mihaela Vorvoreanu. Overreliance on AI Literature Review.

[15] Lucciana Alvarez Ruiz. 2023. Gender Bias in AI: An Experiment with ChatGPT in Financial Inclusion. *Center for Financial Inclusion*. Retrieved February 12, 2026 from https://www.centerforfinancialinclusion.org/gender-bias-in-ai-an-experiment-with-chatgpt-in-financial-inclusion/

[16] Jonathan Sherbino, Kulamakan Kulasegaram, Elizabeth Howey, and Geoffrey Norman. 2014. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: acontrolled trial. *Can. J. Emerg. Med.* 16, 1 (January 2014), 34–40. https://doi.org/10.2310/8000.2013.130860

[17] Philipp Winder, Christian Hildebrand, and Jochen Hartmann. 2025. Biased echoes: Large language models reinforce investment biases and increase portfolio risks of private investors. *PLOS One* 20, 6 (June 2025), e0325459. https://doi.org/10.1371/journal.pone.0325459

[18] Human in the loop: Why human oversight still matters in AI-driven risk and compliance. Retrieved February 12, 2026 from https://www.moodys.com/web/en/us/insights/ai/human-in-the-loop-why-human-oversight-still-matters-in-ai-driven-risk-and-compliance.html

[19] More AI Assistance Reduces Cognitive Engagement: Examining the AI Assistance Dilemma in AI-Supported Note-Taking. https://doi.org/10.1145/3757632

[20] The Principles and Limits of Algorithm-in-the-Loop Decision Making. https://doi.org/10.1145/3359152

A APPENDIX

A.1 FairLend Prototype Links

The following links provide access to the interactive Figma prototypes built for this research.

Condition A — LendCore Standard Dashboard
https://www.figma.com/proto/j4hONhExUEQ4EOqBqMi132/Bias-Busters-%E2%80%94-HCAI-Semester-Project?node-id=2-2&p=f&t=tLzLdgpHShb4goBy-1&scaling=min-zoom&content-scaling=fixed&page-id=0%3A1

Condition B — FairLend Human-Centered Dashboard
https://www.figma.com/proto/j4hONhExUEQ4EOqBqMi132/Bias-Busters-%E2%80%94-HCAI-Semester-Project?node-id=19-301&t=cqfYxGuzlhGe7amy-1&scaling=min-zoom&content-scaling=fixed&page-id=1%3A2&starting-point-node-id=19%3A298

To view the interactive prototype, open the link in a browser. No Figma account is required. Click the play button in the top right corner of the Figma viewer to enter presentation mode and navigate through all four steps.