

# Keeping tabs on the algorithm: How human-AI teamwork can improve loan decisions

By Marshall Terrill , ASU News  
February 26, 2026

As artificial intelligence takes on a larger role in deciding who gets a mortgage, a small-business loan or a line of credit, many banks assume the safest course is to trust the machine.

Algorithms can sift through mountains of data, spot patterns invisible to the human eye and deliver fast, consistent recommendations. In high-stakes financial decisions, that promise of precision is hard to resist.

But new research from Arizona State University suggests that complete trust in AI may not be the smartest strategy. In a real-world field experiment, a researcher in ASU's [W. P. Carey School of Business](#) found that loan officers made more accurate and fairer decisions when they were encouraged to critically evaluate rather than automatically accept AI recommendations.

In some cases, pushing back against the machine led to better outcomes for both lenders and borrowers.

Here, [Department of Information Systems](#) Assistant Professor [Tian Lu](#) discusses his recent paper, "[The Power of Disagreement: A Field Experiment to Investigate Human-Algorithm Collaboration in Loan Evaluations](#)," and explains why a little skepticism may be one of the most powerful tools for responsible AI.

*Note: Answers have been edited for length and/or clarity.*

**Question:** Your study suggests that loan officers make better decisions when they don't automatically agree with AI recommendations. That feels counterintuitive. What's happening psychologically or behaviorally when humans push back against the algorithm?

**Answer:** It may seem counterintuitive because the algorithm — including the one used in our real-world study — was already highly accurate. In most cases, simply following the algorithm was the sensible choice. The performance gains came from a relatively small set of cases in which the algorithm made mistakes and human evaluators corrected those specific errors. The improvement

did not come from frequent disagreement, but from well-targeted correction.

What proved critical was what happened when evaluators encountered a recommendation that conflicted with their judgment, especially when the algorithm's rationale appeared internally inconsistent. When the model's signals didn't seem consistent with the case, evaluators were significantly more likely to disagree. And those disagreements were much more likely to improve the final decision.

Importantly, disagreement itself was not always beneficial. We found that excessive disagreement worsened decision-making. The collaborative value emerged when evaluators overrode the algorithm primarily in cases where they detected algorithmic "self-contradiction." In other words, the behavioral mechanism is not skepticism toward AI, but sensitivity to inconsistency. That selective responsiveness is what allowed the human-AI partnership to outperform either side on its own.

**Q: In your real-world field experiment, how did you encourage "structured disagreement," and what impact did it have on both accuracy and fairness in lending decisions?**

**A:** We varied whether evaluators could see the algorithm's key explanatory signals. When the rationale was disclosed, evaluators were better able to detect cases where the model's highlighted factors appeared internally inconsistent. That transparency did not substantially increase disagreement rates, but it made overrides more selective and more effective.

In terms of outcomes, human-AI teamwork improved decisions by correcting algorithmic mistakes at the margin. More importantly for fairness, the collaboration helped rebalance two types of errors. A Type I error occurs when a borrower is approved but later defaults; a Type II error happens when a qualified borrower is wrongly rejected. Structured disagreement reduced those costly rejections without significantly increasing risky approvals. That shift matters for fairness because wrongly rejecting qualified applicants causes real harm. The gains came not from shifting overall approval standards, but from targeted corrections in conflicted cases.

**Q: Algorithmic bias in lending has been a major concern for regulators and consumers alike. How does your research change the way banks should think about managing bias and accountability in AI-driven decisions?**

**A:** Our research suggests that managing bias in AI-driven lending requires shifting the focus from the model alone to the entire decision system. In practice, lending decisions are produced by a human-algorithm interaction. In our field experiment, outcomes improved only when human intervention was disciplined and grounded in identifiable signals rather than intuition. Design choices such as whether model reasoning is disclosed and how conflicts are surfaced directly shaped whether disagreement strengthened or weakened performance.

This reframes how banks should think about accountability. Bias and error can originate from three sources: the model, the human decision-maker and their interaction. Even a highly accurate algorithm can produce distorted outcomes if overrides are inconsistent or poorly governed. Managing bias, therefore, requires more than model validation. It requires clearly defining who has decision-making authority, setting rules for when AI recommendations can be overridden, and regularly reviewing how those choices affect approval and rejection mistakes. Accountability, in other words, rests with the entire decision-making system — not just the algorithm.

**Q: We often hear that AI reduces human error. Your findings suggest the relationship is more nuanced. What's the right balance between trusting the machine and trusting human judgment?**

**A:** The balance between trusting the machine and trusting human judgment is the balance between reliance and oversight. Algorithms provide consistency and scale. They provide a strong starting point and reduce inconsistency across large volumes of cases. But no model is infallible. Our study shows that performance improves when institutions carefully design interactions rather than leaving them informal. That means making model reasoning visible when appropriate, highlighting signals that help humans spot potential mismatches, and structuring when and how overrides should occur. It also means evaluating outcomes not just by overall accuracy but by how different types of errors shift under collaboration.

So the right balance is not shared authority or constant skepticism. It is a deliberate architecture: The algorithm anchors routine decisions, and human judgment operates within clearly defined protocols that are evidence-based, interpretable and auditable.

**Q: It's Financial Literacy Month and the height of tax season — times when many Americans are thinking about credit, loans and financial stability. What should consumers understand about how AI may be influencing the financial decisions that affect their lives?**

**A:** Today, many financial decisions are shaped by human-algorithm systems. AI models analyze historical patterns to inform loan approvals, credit limits and pricing. They often improve consistency, but they also inherit patterns from past decisions. If historical decisions contained imbalances, those patterns could persist unless they are actively addressed.

For consumers, this means financial literacy now goes beyond understanding interest rates and credit scores. It includes recognizing that repayment history, credit behavior and even digital footprints can shape how automated systems evaluate risk. Making payments on time, keeping credit card balances low and regularly checking your credit report matter even more when AI systems are evaluating your financial profile. Understanding that these systems are based on statistical patterns rather than personal judgment can also help consumers interpret outcomes more rationally.

Equally important is awareness and agency. Consumers should feel empowered to ask how decisions are made, to request explanations when possible and to correct inaccurate records. As AI becomes more embedded in financial systems, understanding how data, models and oversight interact can help individuals navigate credit markets more confidently and proactively.

---

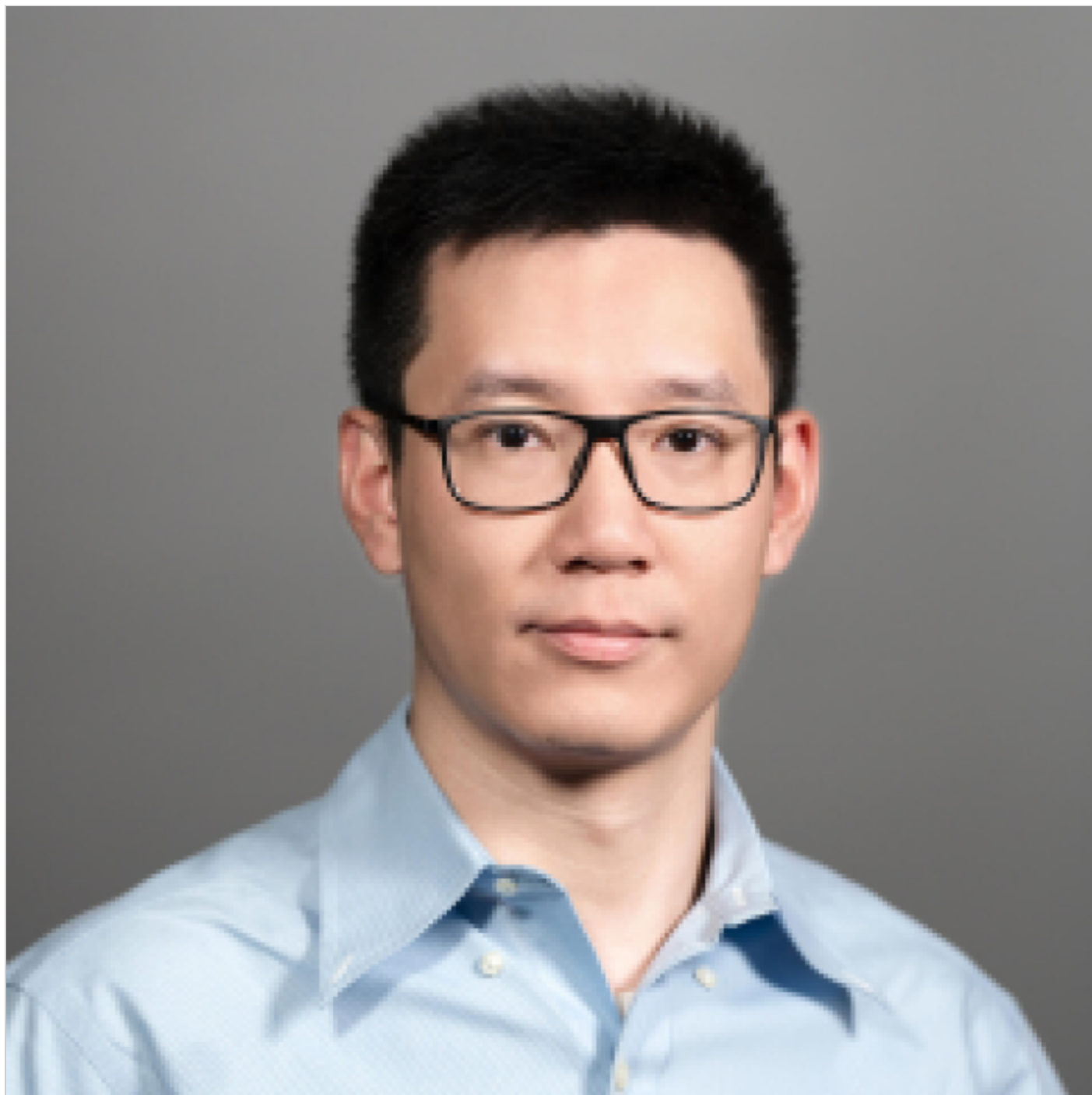
*This story originally appeared on [ASU News](#).*

**Main image**



iStock image

**Text image(s)**



Assistant Professor Tian Lu