Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Ethical Architecture in AI-Driven Credit: Balancing Inclusion, Fairness, and Transparency

Dennis Sebastian
Founder, BINarrator.ai, USA

ARTICLE INFO

Received: 10 Aug 2025

Revised: 15 Sept 2025

Accepted: 24 Sept 2025

ABSTRACT

Artificial Intelligence is now radically transforming credit decisioning systems, enabling unparalleled opportunities for financial inclusion, yet also raising tough implications for bias, discrimination, and transparency. As machine learning algorithms take on more work previously performed by underwriting, financial institutions are now having to confront key trade-offs in predictive accuracy, fairness, and accountability. This article analyzes the social effects of AI-based credit systems through a variety of perspectives, including economic implications, social equity aspects, regulatory evolution, and environmental sustainability. A broad ethical architecture framework is proposed, founded upon four foundational pillars: Inclusive Data practices actively sourcing diverse datasets, Explainable Models that utilize methodologies like SHAP to offer understandable decision rationales, Fair Governance implementing systematic bias detection and audit, and Human Oversight that guarantees expert review of consequential decisions. Real-life case illustrations show both the transformative power of alternative data in widening access to credit for low-income and minority groups and the risks of dark algorithms that embed old discrimination in proxy variables. Regulatory regimes in leading jurisdictions increasingly treat credit scoring as high-risk applications, subjecting them to conformity testing, ongoing monitoring, and thorough impact assessments. The struggle between technological creativity and moral accountability characterizes the present, with organizations facing challenging trade-offs between model performance and interpretability, efficiency and fairness, automation and human judgment. Emerging trends indicate obligatory fairness audits, unified transparency reporting requirements, hybrid human-AI governance mechanisms, and algorithmic impact assessments akin to environmental reviews, reshaping competitive forces in financial services fundamentally towards trustworthiness and social accountability.

Keywords: Algorithmic Fairness, Credit Discrimination, Explainable Artificial Intelligence, Financial Inclusion, Regulatory Compliance

1. Introduction

1.1 Contextual Background

Credit is a pillar of contemporary economies, framing possibilities for individuals and firms. AI-powered underwriting has, in recent years, presented itself as a revolutionary force, using enormous datasets and machine learning methodologies to assess risk faster and, in theory, more accurately than previously possible. The international credit market is being rewritten at its foundation as financial institutions embed machine learning models into their decision-making processes. Traditional credit assessment practices, which for many decades have been based heavily on bureau-based scoring models, are being complemented and even supplanted by models capable of analyzing thousands of data points in milliseconds.

This transition holds significant advantages that reach far beyond operational effectiveness. Alternative sources of data, such as rental payment history, utility bill information, mobile phone use patterns, and

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

education credentials, can bring millions of consumers who are credit invisible or have too little credit history to produce traditional credit scores into the fold. Financial regulatory agency research shows that a large percentage of consumers have credit files too thin to produce traditional credit scores. Automation can accelerate approvals from days to minutes, lower operating costs dramatically in lending operations, and facilitate customized loan offerings that map more accurately to individual risk profiles.

Yet there is another side to this technological revolution. AI models learned on biased historical data have the potential to perpetuate and even to enlarge decades-long inequities built into lending operations. Systematic studies of mortgage solicitations have determined that algorithmic lending systems approved minority mortgage applicants at measurably higher interest rates than those extended to similar white applicants, equaling significant additional fees over the duration of typical mortgages [1]. Left to their own devices, obscure systems can produce discriminatory results that are not apparent to lenders or regulators, which erodes decades of advances in fair lending. The computational overhead of contemporary neural networks, which have millions of parameters trained using gradient descent over billions of examples, generates what researchers describe as the "explainability paradox," wherein the most accurate models are less interpretable.

There is much at stake: credit decisions affect not just financial access but also social mobility, housing prospects, entrepreneurship opportunities, and even intergenerational wealth accumulation. Turned-down mortgage applications can bar homeownership that would have created considerable equity in the long run. Disapproved small business loans can strangle entrepreneurship that could have generated scores of jobs. Credit scoring differences are responsible for wealth disparities where white families possess considerably more wealth than Black and Hispanic families. AI system development must not only be a technical problem solved with sophisticated bias detection tools and fairness metrics—it is a social imperative whose resolution will decide whether artificial intelligence will be a force for democratizing finance or means of encoding discrimination at unheralded scale.

1.2 Problem Statement and Research Gap

Three related issues confront the financial sector that have grown more critical as AI growth accelerates in lending operations. First, bias and discrimination result when historical disparities in credit information infuse structural disadvantages into algorithmic models. Decades of training data used for lending decisions necessarily capture patterns during times when outright redlining practices were legal, when gender-based credit discrimination was the norm, and when systemic barriers restricted credit access to protected classes. When the machine learning algorithms maximize patterns within the data, they will learn to simulate discriminatory results even if protected attributes like race and gender are manually left out of model inputs. It has been shown through research that neutral-looking variables like zip codes, patterns in employment history, and some demographic proxies can reinstate historical prejudices with statistically significant correlations.

Second, many AI models are opaque in that they are "black boxes" that regulators and consumers cannot comprehend. Contemporary deep learning architectures, such as gradient boosted decision trees with many weak learners and neural networks with several hidden layers, can make precise predictions but give little insight into the justification for specific decisions. Surveys of compliance officers in large financial institutions indicate widespread worry over their inability to provide explanations of adverse action reasons when automated systems make credit decisions, and potential violations of rules that demand clear explanations for credit denials. This is compounded by ensemble techniques that use ensemble outputs from multiple models, proprietary algorithms shielded as trade secrets, and continuously updated models through online learning processes.

Third, there is regulatory uncertainty because the legal frameworks have not yet fully embraced the sophistication of algorithmic credit decisioning. The underlying consumer protection laws were written

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

for a time of human underwriters and low-dimensional scorecards, not of machine learning algorithms that can detect subtle nonlinear relationships among hundreds of variables [2]. Regulators in Europe have taken further steps with holistic AI regulation, designating credit scoring systems as high-risk applications subject to conformity assessments, although guidelines for implementation are meager. Other regulatory bodies have provided interpretive guidance on algorithmic decision-making, but uncertainties remain regarding how existing adverse action notice obligations apply where models are incapable of producing human-interpretable explanations.

As institutions try out AI, implementing systems across consumer lending, small business finance, and mortgage origination, they lack codified ethical architecture frameworks in financial credit systems. The available technology company and academic institution frameworks discuss AI ethics in general terms but do not include the particular regulatory requirements, fairness definitions, and risk management practices peculiar to financial services. This disconnect generates inconsistent execution wherein every institution forges its own bespoke methods without normalized benchmarks, independent verification practices, or industry-best guidelines for assessing and reducing algorithmic bias in credit decisioning.

1.3 Purpose and Scope

This paper considers the ethical threats and benefits of AI in credit by a multidisciplinary framework that brings together technical machine learning considerations, regulatory compliance mandates, and social fairness imperatives. It offers a four-pillar structure for fair system design based on both theoretical principles of fairness and practical limits of financial institution implementation. The structure integrates findings from computer science research on algorithmic fairness, legal research on anti-discrimination law, and empirical research on implemented credit systems to offer practical recommendations for institutions looking to benefit from AI's powers while limiting its dangers.

The paper reviews actual case studies of inclusion success and bias failure across both, using documented events, regulatory enforcement actions, and peer-reviewed analyses to demonstrate the tangible expressions of ethical design decisions. Such cases involve alternative lending websites that have increased credit availability to traditionally underserved groups, technology-mediated discrimination that went undetected by traditional monitoring, and regulatory actions that have influenced industry norms. By comparing failures with successes, the paper uncovers typical pitfalls in AI credit system design and signals precautionary measures that work in real life.

The debate encompasses long-term social and regulatory consequences well beyond specific lending choices to discuss system effects on wealth disparities, inclusion in finance, and economic mobility. It considers how AI-based credit systems engage with current social structures that may support or challenge established mechanisms of resource allocation. This article looks at the development of new regulatory regimes in several jurisdictions and how shifting legal standards will impact the development and deployment of credit AI systems in the coming years. Finally, it explores developments in both explainable AI, fairness-sensitive machine learning, and human-AI collaboration that can mitigate current trade-offs between model performance and ethical demands, establishing a framework for knowledge-based innovation in this important space.

1.4 Statistics Relevant to the Topic

Recent studies illustrate the scale of such challenges by using empirical metrics of AI credit system performance, expressions of bias, and industry patterns of adoption. Alternative lending platforms based on AI and alternative data sources have shown quantifiable gains in credit availability among populations not well served by traditional scoring techniques, while having similar default rates for their portfolios of loans. Machine learning algorithms using large alternative variables such as education credentials, work history, and niche data points have made it possible the approve more borrowers each year who would

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

have been rejected under legacy criteria, with special success in opening up access to younger borrowers whose thin credit files contain inadequate information for standard scoring.

But ongoing disparities in legacy credit metrics underscore the size of past inequalities that AI systems will need to contend with. Research on the credit score distribution of demographic groups reveals quantifiable disparities between minority and white lenders, with lower average scores for Black and Hispanic homebuyers. These disparities indicate compounding disadvantages from intergenerational wealth disparities, as white families maintain significantly higher levels of wealth than Black and Hispanic families. Lower credit scores directly correlate to greater cost of borrowing, with low-score range borrowers paying mortgage interest rates far exceeding their top-score counterparts, adding up to huge additional interest payments over common mortgage lifetimes [1].

Implementation issues pose great operational obstacles to fair AI deployment in credit decisioning. Extensive surveys of financial services business leaders reflect that explainability and fairness are leading issues in AI adoption over technical performance or integration price concerns. Within institutions that have implemented AI credit models, high percentages of them report challenges in explaining unfavorable actions to consumers in terms that meet regulatory standards, while others have a hard time performing meaningful bias audits in the absence of standardized fairness metrics. The technical intricacies of advanced models exacerbate these challenges, with top credit AI models utilizing gradient boosted decision forest ensembles with hundreds of trees of considerable depth, making human interpretation effectively impossible.

Regulatory drive toward compulsory transparency and responsibility continues to gather pace in key jurisdictions. Thorough AI regulation in key economic blocs categorizes credit scoring systems as high-risk applications subject to conformity testing, monitoring, and extensive technical documentation before deployment. In accordance with these stipulations, thousands of AI credit systems rolled out across various jurisdictions would need to be independently validated and have audit trails that are explainable for individual decisions. Regulatory agencies have made supervisory statements highlighting that using AI does not relieve institutions of requirements to give notice of adverse action under consumer protection laws, which can impact thousands of depository institutions and nonbank lenders running AI-based underwriting systems. Regulatory agencies have initiated multi-year projects studying algorithmic discrimination in consumer finance, and early results have suggested that large percentages of AI credit models explored had statistically significant differences in approval rates across protected demographic categories, even after accounting for credit risk characteristics [2].

2. Impact Analysis and Ethical Framework

2.1 Wider Impact

Artificial intelligence-based credit models can democratize finance by using alternative data sets—such as utility bills, mobile transactions, rental histories, and educational attainment—to judge borrowers with few standard credit histories. Next-generation machine learning platforms have shown their potential to revolutionize the extension of financial inclusion to historically underserved groups of people. Alternative lending sites that use neural networks and ensemble techniques have realized quantifiable approval rate gains for thin-file borrowers while preserving portfolio performance metrics in line with those using traditional underwriting strategies. These sites evaluate vast alternative points of information, such as regular bill payment history, educational attainment, employment stability factors, and transaction behavior that are associated with creditworthiness but are not visible to traditional scoring systems.

The democratizing power is especially applicable to younger consumers, new immigrants, and those recovering from financial losses who have legitimate repayment ability but not the long credit histories demanded by legacy models. Machine learning models can see predictive trends in rent payment

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

reliability, utility account maintenance, and telecom billing that reflect financial responsibility even without thin bureau files. Alternative data sources allow institutions to shift away from bureau-based judgments, disproportionately excluding populations with no history of engagement with legacy credit products. Additionally, alternative data incorporation diminishes dependency on intergenerational wealth surrogates that unintentionally entrench prior disadvantages, opening doors to credit for populations whose parents did not have access to mortgage funding or business loans.

However, risks are deep and increasingly established via regulatory examinations and empirical studies. High-profile algorithmic bias research has identified that even when algorithms are not necessarily employing protected characteristics like gender, proxies built into training data can establish systemic disadvantages for women and minority groups. Financial regulators who undertook in-depth investigations found that apparently innocuous variables such as patterns of transaction categorization, measures of account balance volatility, and patterns of spending categories were able to act as demographic proxies, leading to measurably varying credit limits and interest rates for applicants sharing the same risk profile [3]. The proxy methods work through sophisticated correlation patterns that evade traditional bias detection techniques designed to be attentive only to direct use of protected attributes.

In parallel, encoded demographic patterns research illustrates how seemingly minor correlations in monetary data can penalize particular groups even as there exists empirical evidence of similar or superior repayment behavior across various categories of lending. Academic studies comparing large credit account portfolios have reported that demographic cohorts presenting lower rates of default and more consistent payment habits can still be assessed at systematically lower credit when algorithmic models learned based on past data are fed in different data sources. The mechanism works by variables that are associated with demographic traits—like patterns of income stability, categorizations of spending, and profiles of employment history—that algorithms read as risk factors even though they have weak predictive abilities for real default behavior.

The historical biases that are amplified by algorithms pose especially pernicious issues because the discrimination is working through statistically sound correlations in training data rather than overt prejudicial rules. Machine learning algorithms maximizing predictive performance tend to lean towards patterns that mirror past lending results, which themselves contain decades-worth of institutional bias [4]. During the era when credit decisions were made primarily manually, discriminatory behavior openly rejected applications from well-qualified applicants who belonged to protected groups. Contemporary algorithms relearn these patterns computationally, applying bias at scale without knowing it. The outcome is algorithmic discrimination that is hard to identify using standard auditing, since the models obtain excellent accuracy measures while consistently disadvantaging protected groups indirectly through mechanisms hidden from standard fairness audits.

2.2 Responsibility and Equity

Legal structures like the Equal Credit Opportunity Act in the US and the Consumer Credit Directive in the EU require fairness and lack of discrimination in credit determination, setting guiding principles that are pre-algorithmic in origin but still entirely applicable to AI-based systems. These laws prevent discrimination on protected grounds such as race, color, religion, national origin, sex, marital status, and age, mandating that credit decisions be based on factors demonstrably relevant to creditworthiness only. The regulatory design goes beyond mere prohibition and mandates positive duties of transparency such that lenders must provide detailed reasons for adverse actions and keep records adequate to prove compliance with anti-discrimination requirements.

Ethical design should thus incorporate four core elements that operationalize legal mandates within algorithmic systems. First, equitable data practices actively procure diverse datasets to preclude exclusion and rigorously assess whether alternative data sources bring new types of bias. Financial institutions

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

using responsible AI frameworks make systematic surveys of training data populations, assessing representation across protected groups and determining variables that can act as proxies for barred characteristics [3]. Such analyses move beyond basic demographic balance to analyze intersectional impacts where sets of characteristics combine to produce special disadvantages. Data sourcing approaches specifically include data from a variety of geographic areas, economic segments, and demographic categories so models learn from representative populations, not historically advantaged groups.

Second, explainable models give transparent reasons for choices with methods like Shapley Additive explanations, Local Interpretable Model-agnostic Explanations, and counterfactual reasoning systems that determine what factors contributed to a given outcome. SHAP values, based on cooperative game theory, assign each of an input feature's contributions to a prediction by calculating marginal contributions over all feature combination possibilities, delivering mathematically sound explanations even for sophisticated ensemble models. Deployment of explainable AI methods allows institutions to produce adverse action notices that comply with regulatory mandates by determining the specific factors—like debt-to-income ratio, payment history discrepancies, or credit usage patterns—that most affect denial decisions.

The explainability requirement is not limited to individual decisions to model-level interpretability, in which institutions must know how algorithms weigh factors systematically. Sophisticated interpretability models produce global explanations displaying rankings of feature importance, partial dependence plots of how the predictions vary with variable ranges, and interaction effects describing how sets of factors impact outcomes [4]. Such tools allow compliance teams to ensure that models are based on genuine creditworthiness factors and not surrogates for protected attributes, facilitating regulatory compliance as well as ethical stewardship.

Third, equitable governance sets up fairness audits, bias detection pipelines, and regulatory reporting procedures as part of the model development life cycle. Governance frameworks have multi-step review processes such that models are subject to bias testing before deployment, ongoing monitoring during production use, and occasional end-to-end audits measuring cumulative impacts. Fairness audits utilize several statistical concepts of fairness—demographic parity, equalized odds, and calibration equity—since various fairness notions capture different facets of discrimination and, in fact, might contradict each other in practice. Bias detection pipelines automatically mark models showing statistically significant differences between protected groups, initiating human scrutiny and possible model tuning prior to adverse impacts adding up.

Fourth, human supervision guarantees that human professionals to examine adverse actions and watch over systemic effects, keeping in place oversight mechanisms that ensure that consequential decisions are not left entirely to machines. Oversight frameworks place seasoned credit analysts in a position to review algorithmically flagged marginal instances, examine samples of automated denials for conformity with institutional procedures, and probe anomalous patterns indicative of emergent bias. Human evaluators use contextual judgment that cannot be emulated by algorithms, taking into account unusual situations, assessing explanation quality from the perspective of the applicant, and recognizing points at which statistical prediction may not fully measure individual creditworthiness. This human-in-the-loop design weighs efficiency gains through automation against the irreducible requirement for human judgment in high-risk financial choices impacting individuals' economic prospects.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Pillar	Key Components	Implementation Mechanisms
Inclusive Data	Diverse dataset sourcing, demographic representation measurement, proxy variable identification	Systematic inventory of training data, geographic and economic diversity incorporation, intersectional effect assessment
Explainable Models	SHAP values, LIME techniques, counterfactual reasoning frameworks	Feature importance rankings, partial dependence plots, adverse action notice generation
Fair Governance	Fairness audits, bias detection pipelines, regulatory reporting	Multi-stage review processes, demographic parity testing, equalized odds measurement
Human Oversight	Expert review of adverse actions, borderline case examination, systemic pattern investigation	Human-in-the-loop architecture, contextual judgment application, algorithmic override pathways

Table 1: Four-Pillar Ethical Architecture Framework for AI Credit Systems [3, 4]

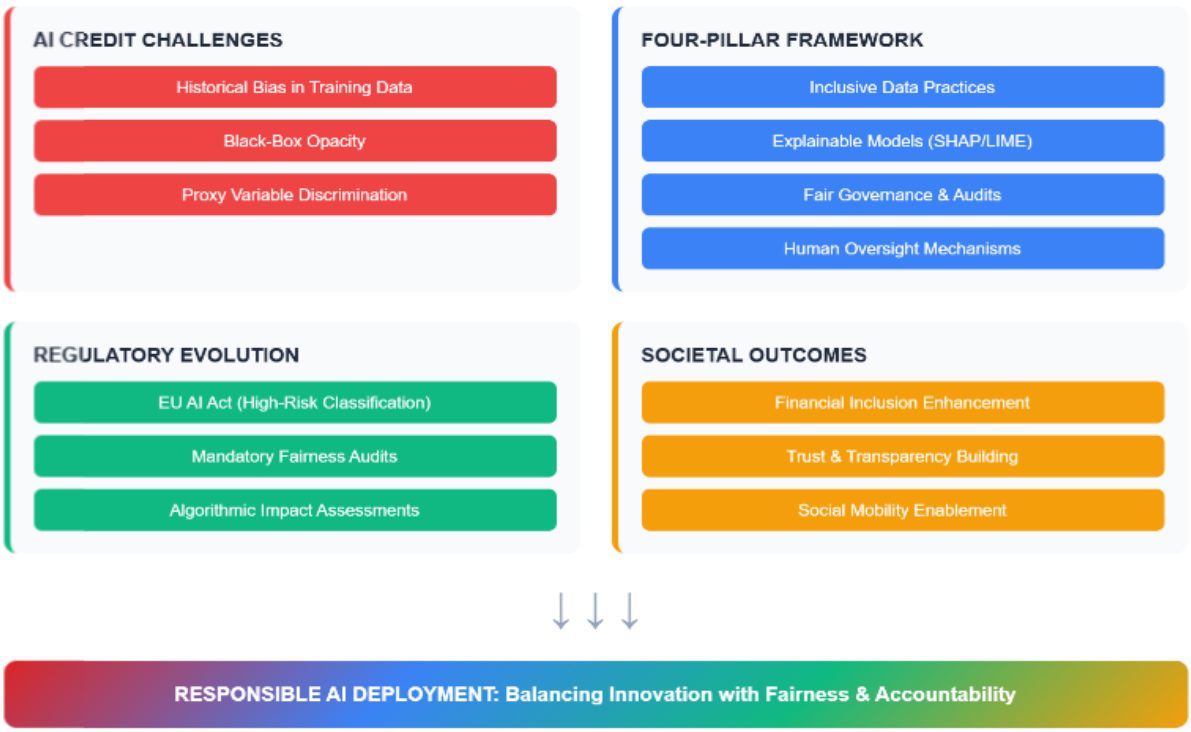

Fig. 1: Ethical Architecture in AI-Driven Credit: Balancing Inclusion, Fairness, and Transparency.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. Policy, Regulation, and Case Studies

3.1 Policy and Regulation Impact

Regulators are growingly concerned with AI transparency as algorithmic systems of decision-making expand throughout consumer financial services. Various regulatory bodies have issued guidance, enforcement actions, and formal regulations addressing the distinctive problems raised by machine learning models in credit underwriting. The U.S. Consumer Financial Protection Bureau cautioned against inscrutable algorithms that don't give good reasons for taking unfavorable actions, noting that technical savvy doesn't relieve institutions from long-standing fair lending requirements in the Equal Credit Opportunity Act and Fair Credit Reporting Act. The Bureau has indicated increased supervisory examination for institutions using AI systems, performing focused examinations of algorithmic underwriting systems, and issuing interpretive guidance making clear that adverse action notice rules apply whether or not the decisions arise from human consideration or automation.

The European Union's AI Act is the most far-reaching regulatory environment for algorithmic systems in the world, creating a risk-based categorization system dividing applications based on their possible societal effect. The legislation classifies credit scoring and creditworthiness assessment systems as high-risk applications subject to stringent requirements before market deployment. High-risk credit models must undergo explainability assessments demonstrating that institutions can provide clear, specific reasons for individual decisions that affected persons can understand and contest. Bias evaluations involve systematic examination of model performance by demographic group, with statistical testing for disparate impact and documentation of mitigation steps when disparities are found [5].

The regulation framework requires conformity assessment procedures involving independent third-party auditing before deployment for some systems with high risk. Continuing monitoring responsibilities charge institutions with putting in place quality management systems that constantly monitor model performance, identify drift in prediction patterns, and reveal signs of emerging bias indicators during the system life cycle. Documentation requirements list that institutions must keep complete technical records about training data sources, model architectures, validation methods, and fairness testing outcomes. Regulatory reporting frameworks involve regular reporting of performance statistics to regulatory authorities, facilitating early intervention in algorithmic systems implemented at scale throughout financial markets.

The UK Financial Conduct Authority has emphasized digital lending fairness as a cross-cutting strategic priority within its overall consumer protection remit, commencing multi-year work on algorithmic bias in retail financial services. The Authority has released discussion papers setting out expectations for treating customers fairly when rolling out AI systems, stressing that algorithmic decision-making does not diminish institutional responsibility for discriminatory treatment. Supervisory strategies use thematic reviews of algorithmic behavior across a set of institutions alongside focused investigations of individual platforms following consumer concern or market intelligence indicating the possibility of bias. The Authority has indicated a willingness to exercise enforcement powers on institutions whose algorithmic systems deliver discriminatory results, whether resulting from deliberate design or emergent model behavior.

Regulatory convergence across borders is an indication of increasing international agreement that algorithmic credit systems need oversight regimes tuned to their distinctive nature [5]. Global standard-setting organizations have established guideline papers on AI governance, model risk management, and fairness test approaches specific to financial services environments. These standards highlight principles such as transparency in model creation, accountability for algorithmic results, fairness testing across the model life cycle, and human oversight of consequential decisions. Implementation schedules differ

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

between jurisdictions, with some requiring institutions to be compliant immediately and others providing transition periods within which institutions can adjust existing systems to new standards.

3.2 Case Examples

There are a number of cases that reflect the dual character of AI in credit, showing both its transformative power for financial inclusion and huge potential risks of encoded discrimination. Online alternative lending platforms that leverage machine learning and alternative data have shown documented success in extending access to credit to underserved populations beyond traditional scoring models. Sophisticated algorithmic systems have approved many more applicants from underserved population groups with portfolio loss rates commensurate with traditional underwriting strategies. These platforms show that non-traditional sources of data, such as rental payment history, utility account management, educational verification, and employment stability trends, can both accurately forecast creditworthiness for those who do not have traditional credit files. The inclusion success is attributed to models' capacity to detect predictive patterns obscured by bureau-based scoring, allowing extension of credit to deserving borrowers who would otherwise be subject to systematic denial under conventional methodologies.

Yet, headline-grabbing bias incidents highlight the risks of hidden systems implemented without sufficient fairness protection. Regulatory probes into large credit card schemes have uncovered systemic gender bias in which women applicants were being offered significantly lower credit lines than men with similar or better financial histories. The study, performed by state finance regulators after public complaints were made, reviewed vast credit decisioning files and found statistically significant gender gaps that were not explainable by valid risk factors. Higher-income female applicants with better credit scores and lower debt-to-income ratios consistently were issued substantially lower credit limits than similar male applicants. The algorithmic system, although not directly accounting for gender as an input variable, was trained to mimic discriminatory patterns on proxy variables linked to gender, such as spending category distribution, account balance patterns, and transaction attributes.

The case revealed inherent difficulties in detecting and preventing algorithmic bias. The credit card company asserted that its algorithm did not take gender into account and thus could not discriminate, an overly narrow technical concept of fairness, which regulatory probes demonstrated was insufficient. The algorithm had found correlations between valid financial variables and gender and was employing these correlations to apply discrimination indirectly in order to have high predictive accuracy measures [6]. Conventional fairness audits, checking only overt use of covered attributes, did not find the bias, and it took advanced disparate impact analysis comparing results across demographic groups to find the systematic disadvantage.

Studies show that mixes of age, gender, and parental status can produce complex, cumulative disadvantages through intersectional discrimination that demographic parity analysis missed. Algorithmic systems can be fair to each protected attribute individually when designing, but impose extreme disadvantages on those with multiple protected characteristics at the same time. Research examining credit decisioning algorithms has found that some demographic subgroups experience disproportionate denial relative to additive predictions of individual attributes. The resulting compounded disadvantage arises from interaction effects where algorithms discover that combinations of characteristics produce different predictions than individual attributes examined separately.

Intersectional discrimination is especially challenging to identify because standard fairness audits test protected attributes in isolation instead of looking at the treatment of subgroups delineated by more than one characteristic. An algorithm can show demographic parity for the protected group as a whole, but prejudicially treat certain subgroups delineated by more than one characteristic. Mathematical complexity ramps up exponentially as more characteristics are taken into account, with full intersectional evaluation being necessary to analyze many possible combinations of attributes [6]. This computational and

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

conceptual problem has compelled many institutions to adopt incomplete fairness tests that overlook intersectional harms, sustaining discriminatory results that are not detectable with typical audit procedures. Mitigating intersectional discrimination calls for nuanced analytical tools that consider subgroup treatment in explicit form, going beyond aggregate measures of fairness to assess outcomes for groups defined by multiple protected aspects jointly.

Jurisdiction	Primary Regulatory Approach	Key Requirements
European Union	Risk-based classification under AI Act	Conformity assessments, explainability demonstrations, bias testing, independent third-party auditing
United States	Compliance with existing fair lending statutes	Adverse action notice requirements, CFPB supervisory scrutiny, interpretive guidance on algorithmic decisioning
United Kingdom	Strategic priority within consumer protection mandate	Thematic reviews across institutions, discussion papers on fair treatment, enforcement against discriminatory outcomes
International Bodies	Consensus on adapted oversight frameworks	AI governance guidance, model risk management protocols, fairness testing methodologies

Table 2: Regulatory Framework Comparison for AI Credit Systems [5]

4. Broader Implications

4.1 Economic Implications

Ethical AI opens credit markets by bringing into view hitherto excluded borrowers, reshaping the composition and size of consumer and commercial lending. Machine learning models with non-traditional data sources allow financial institutions to provide credit to hitherto excluded populations, with significant economic multiplier impacts. When well-screened borrowers can access credit that would otherwise be unobtainable, they can fund education, buy homes, start businesses, and smooth consumption in times of income disruption. These actions create positive externalities across the economy as homeownership fosters wealth and neighborhood resilience, investments in education enhance productivity, and business creation generates jobs.

The economic growth from inclusive credit accrues particularly to underserved geographic areas and demographic segments. Rural villages without physical bank locations are able to use digital lending

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

platforms with algorithmic underwriting based on alternative data. Young adults who are entering the workforce can receive their initial credit products without needing extensive payment histories that legacy models require. Immigrants starting new lives in new countries can establish creditworthiness through rental payments and utility bills instead of waiting long periods to develop traditional credit files. Each new borrower added to the pool is an additional economic activity, with credit facilitating purchases and investments that otherwise would remain delayed perpetually [7].

In contrast, discriminatory systems can exacerbate defaults, perpetuate inequality, and compromise financial stability through several reinforcing dynamics. When algorithmic models consistently misprice risk for specific demographic segments, overestimating or underestimating default probabilities, they introduce portfolio vulnerabilities that add up throughout the financial system. Overestimation results in qualified borrowers being denied or charged excessive interest rates, decreasing the availability of credit and limiting economic activity. Underestimation results in improper extension of credit to borrowers who have no repayment capacity and leads to losses institutions need to incur, and potentially causes wider financial instability if concentration is in particular market segments.

Inequality-entrenching mechanisms of distorted credit systems work through wealth accumulation channels. Mortgage denial prevents homeownership that would create equity appreciation over long time spans. Increased interest rates result in higher payments that otherwise would be used to save for retirement or college investment. Denials of business loans inhibit entrepreneurship that would create jobs and business equity. Such impacts increase over time and generations, since wealth-building parents cannot afford to give down payments, educational assistance, or inheritances to children. The resulting intergenerational transfer of disadvantage reinforces economic stratification along demographic lines, followed by algorithmic bias.

Financial stability issues arise when widespread algorithmic bias produces correlated exposures within institutions. If various lenders use the same machine learning architectures trained on similar data sets, they will reproduce the same biases, consistently overexposing some market segments and underserving others. This correlation produces systemic risks where shocks to the mispriced segments transmit across institutions at the same time. Regulatory bodies overseeing financial stability increasingly appreciate that algorithmic credit decisioning creates new types of correlated risk subject to macroprudential oversight in addition to conventional microprudential supervision [7].

4.2 Social Implications

Transparency, equity, and access in credit systems foster greater trust in financial institutions and facilitate greater social mobility through several interlinked channels. As people comprehend how credit is decided and find the process to be fair, they uphold faith in financial institutions as legitimate arbiters of economic opportunity. Such faith is crucial for financial system stability since banking intrinsically relies upon public trust that institutions will act fairly towards customers and keep promises. Transparent algorithms that explain clear reasons for decisions, even negative ones, enable applicants to identify particular deficiencies and make corrective improvements. This feedback system converts credit denial from a black box obstacle into usable advice for credit development.

Equitable credit systems enable social mobility by preventing economic progress from resting on demographic traits, but instead on individual effort and ability. As qualified borrowers from diverse backgrounds have access to credit on equal terms, education and entrepreneurship are sustainable options for economic advancement regardless of family fortunes or social connections. Access to credit allows gifted individuals with disadvantages to invest in human capital through education, geographic mobility for access to employment opportunities, and business creation. These investments pay returns that snowball over lifetimes and careers, facilitating upward mobility across socioeconomic groups.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The social cohesion gains of equitable credit include community-level impacts. As lending organizations treat neighborhoods fairly, communities gain access to capital for improvements to housing, business growth, and the development of infrastructure. Fair access to credit avoids the redlining cycles that used to target areas of disinvestment in minority areas, generating neighborhood deterioration and wealth destruction. Equitable algorithmic underwriting can actively combat such past trends by assessing borrowers and properties against actual risk factors instead of demographic surrogates that reinforce earlier discrimination.

Alternatively, opacity and bias undermine public trust and strengthen prevailing inequities through mechanisms that are socially corrosive. When credit determinations result from impenetrable algorithms that applicants cannot review or appeal, the process feels arbitrary rather than meritocratic. Such arbitrariness destroys social contract expectations that effort and responsibility will translate into opportunity. Populations subject to algorithmic systematic bias grow legitimate skepticism towards financial institutions, perceiving them as discriminatory gatekeepers rather than as neutral intermediaries.

The equity-reinforcing impact of discriminatory credit systems works through various interactive and cumulative channels. Credit rejection inhibits wealth accumulation that would make future generations able to secure more favorable education, housing, and business opportunities. Increased interest rates drain resources from previously disadvantaged groups and transfer wealth to the advantaged group that possesses financial capital. Geographic credit rejection produces neighborhood disinvestment that lowers property values, school quality, and economic opportunities. These mechanisms build self-reinforcing loops in which original disadvantages build upon themselves over time and distance, deepening inequality by demographics.

Social fragmentation arises when credit systems deliver results that are seen as discriminatory. Populations subjected to algorithmic bias may politically mobilize against financial institutions and their supportive policymakers, leading to regulatory backlash and social strife. Coverage of high-profile cases of bias in the media incites public anger beyond those directly exposed to discriminatory outcomes to larger populations with concerns about fairness. This social tension makes effective policy conversation about financial innovation more difficult, as valid issues around discrimination are wrapped up in more general technological worries and political polarization.

4.3 Environmental Implications

Responsible, cloud-native AI systems decrease the need for manual processes and paper-intensive workflows, decreasing operational overhead and enabling sustainability targets through several avenues. Conventional credit underwriting requires extensive paper documentation, such as application forms, income verification materials, asset statements, and supporting documentation that borrowers need to gather and present. Institutions have physical files during the loan duration, with the need for storage facilities with climate control, security, and administrative personnel. Document shipment between offices, processing facilities, and regulatory records produces emissions from fleets of automobiles. This paper-based infrastructure has a great deal of environmental impact from wood harvesting for paper manufacturing, production, and printing operations, physical storage needs, and final document destruction.

Cloud-based AI credit platforms cut most of this physical infrastructure by going digital for the entire underwriting process. Candidates send documents electronically via web interfaces or mobile apps, dispensing with printing and shipping needs. Machine learning algorithms consume financial institution digital data, utility company data, and other data sources directly through application programming interfaces, minimizing the need for manual documentation. Electronic storage supplants physical file systems, with cloud data centers providing energy efficiency far above that of distributed on-premises

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

storage. The concentration of computing power in large data centers allows for optimization mechanisms such as server virtualization, dynamic workload scheduling, and waste heat recycling that become physically unfeasible with distributed infrastructure.

Yet the environmental mathematics of AI credit schemes continue to be complicated and debated. Training machine learning models takes a lot of energy, especially deep neural networks involving iterative optimization over massive data. Training large ensemble systems can involve computational resources and the resultant carbon footprint. Model retraining to keep pace with changing data distributions places constant energy requirements. AI system proliferation in financial services and other areas fuels data center growth that, while reducing energy intensity per calculation, could result in higher absolute energy usage due to increasing use [8].

Studies considering the climate footprint of information and communication technologies indicate that AI and machine learning make significant contributions to total digital industry emissions. The training energy intensity of big models, along with inference-scale computational needs across millions of transactions, results in environmental profiles that need to be carefully examined. Data center power usage keeps increasing worldwide, fueled in part by growing AI workloads across industry verticals such as financial services. While individual data centers realize better energy efficiency by leveraging technological progress, the total growth in digital infrastructure can nullify these improvements through rebound effects, where decreased costs facilitate increased use.

Financial institutions increasingly factor environmental concerns into AI system design and deployment choices. Sustainable development practices involve model architecture choices in favor of efficient ones, optimization of training schedules to leverage renewable energy availability, and model compression methods that lower inference computational demands. Some organizations perform environmental impact analyses for large AI projects, quantifying carbon footprints and setting reduction targets. Regulatory systems in some locales start to include requirements for sustainability reporting from financial services technology infrastructure, generating transparency regarding the environmental expense of algorithmic systems in addition to their economic and social consequences [8].

Conflict between AI's productivity advantages and energy requirements is an expression of wider dilemmas in technology-facilitated sustainability transformations. Although algorithmic credit systems minimize paper usage and physical infrastructure needs, they transfer environmental impacts to digital infrastructure with associated carbon costs. The overall environmental impact varies with variables such as energy sources for data centers, computational cost of models used, extent of system utilization, and induced demand impacts. Systematic lifecycle analysis becomes a requirement to assess whether certain AI applications translate into real environmental advantages or move impacts around various categories.

Dimension	Positive Impacts	Negative Risks
Economic	Credit market expansion, financial inclusion for invisible borrowers, economic multiplier effects, and entrepreneurship enablement	Systematic risk mispricing, portfolio vulnerabilities, wealth accumulation prevention, and intergenerational disadvantage transmission
Social	Enhanced institutional trust, social mobility support, community capital access, and redlining pattern counteraction	Public confidence erosion, mistrust development, neighborhood disinvestment concentration, and social fragmentation
Environmental	Paper workflow elimination, digital documentation adoption, cloud efficiency optimization, and	Model training energy consumption, data center expansion, aggregate infrastructure growth, and rebound

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	physical infrastructure reduction	effect challenges
Financial Stability	Alternative data utilization, risk assessment improvement, and digital platform accessibility	Correlated institutional exposures, systemic vulnerability creation, and bias-driven market segment concentration

Table 3: Multidimensional Implications of AI-Driven Credit Systems [7, 8]

5. Future Directions and Strategic Recommendations

5.1 Future Outlook

The credit system design ethos will shift towards fairness audits as formal regulatory requirements incorporated into institutional governance systems, with the same standards of scrutiny accorded to financial audits and risk management analysis. Regulators in key jurisdictions are creating standardized fairness testing procedures that will impose pre-deployment verification, ongoing monitoring, and regular, thorough reviews of algorithmic credit systems. The procedures will necessitate institutions to capture outcomes across protected demographic classes, subject these to disparate impact tests using proven statistical techniques, and record mitigation strategies where bias indicators pass threshold levels. The legalization of fairness auditing is a paradigm shift away from self-regulatory ethical pledges to compliance with regulatory oversight and possible penalties.

Hybrid human-AI governance models will bake in checks through workflow designs that place human judgment at key decision-making points instead of relegating humans to post-hoc review. These models of governance acknowledge that good control involves integration into the decision-making process itself and not auditing of already made decisions. Implementation strategies involve human checking of borderline cases flagged algorithmically where model confidence dips below thresholds set, compulsory human authorization for decisions made relating to protected characteristics or sensitive situations, and expert panels reviewing systemic patterns within automatic decisions. The hybrid structure reconciles efficiency benefits of automation with irreducible requirements for human judgment in high-risk decisions impacting individuals' economic prospects [9].

Sociotechnical analysis, however, exposes essential limitations in purely technical fairness solutions. Algorithmic systems function within larger social environments where abstract fairness measures may not capture real-world discrimination processes. Technical interventions taming statistical imbalances at a level of abstraction can inadvertently produce or reinforce unfairness at different levels. For example, maintaining demographic balance in approval rates over a single protected characteristic might hide intersectional discrimination on subgroups defined by multiple attributes. In the same way, optimizing for individual fairness defined through similarity metrics might bake societal prejudices into similarity definitions themselves, reproducing discrimination through seemingly value-free mathematical abstractions.

The challenge involves not only choosing suitable fairness metrics but also the whole sociotechnical system from data collection processes to institutional practices, regulatory regimes, and societal contexts. Credit decisions do not come from algorithmic calculations alone but from intricate interactions among technical systems, human agents, organizational forms, and social institutions. Confronting algorithmic bias demands interventions across various system levels instead of stand-alone technical solutions. Fairness frameworks need to take into consideration how abstractions made at system design contain normative assumptions that benefit some groups but harm others, usually in technically imperceptible ways, and aim only at narrow optimization goals [9].

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Algorithmic impact assessments similar to environmental impact reviews will be legally required for high-risk uses, demanding institutions to make thorough assessments before releasing credit AI systems and regularly afterward. These evaluations will test effects on several dimensions such as fairness across groups, risk prediction accuracy and calibration, resistance to input perturbations and adversarial attacks, privacy aspects of data processing and collection, and systemic influences on credit market dynamics. Impact assessment models will involve quantitative analysis of outcome differentials, qualitative analysis of suspected discrimination mechanisms, stakeholder engagement with impacted communities, and risk mitigation planning. Regulatory direction more and more dictates methodological demands on impact assessments, standardizing measurements, testing practices, and documentation structures. Standardization allows for comparable assessment across institutions and longitudinal monitoring of industry progress toward fairness goals.

5.2 Long-Term Perspective

The decade ahead will find regulators calling for fairness metrics with the same level of detail and standardization that is applied to financial reporting requirements. Regulatory bodies are creating taxonomies of fairness definitions relevant to credit settings, taking into account that several statistical notions reflect varying facets of discrimination. Institutions will present metrics such as demographic parity in approval rates, equalized odds assessing false positive and false negative rate equality, calibration equity guaranteeing predicted probabilities corresponding to actual outcomes by groups, and counterfactual fairness to assess whether decisions would differ if protected characteristics were different. Standardized reporting structures will allow regulators to compare institutional performance, spot outliers that need additional oversight, and monitor sector-wide trends in algorithmic fairness.

Institutions that release transparency reports will make full disclosures regarding AI system design, training data properties, validation methods, and performance metrics by demographic factors. These reports will detail model types implemented, feature sets used, data sources and preprocessing steps, fairness test outcomes, and remediation steps that mitigate detected biases. Transparency reporting will go beyond technical information to include governance arrangements, human oversight processes, consumer complaint channels, and adverse action explanation processes. Public disclosure generates accountability mechanisms that facilitate external scrutiny of institutional practices and competitive pressure towards higher standards of fairness.

The trend toward continuous fairness documentation acknowledges that algorithmic systems must be monitored continuously instead of being certified once. Machine learning models are subject to performance drift as the distribution of input data changes over time, which can introduce emergent biases not present at the time of first deployment. Model revision by retraining or architectural changes can inadvertently cause new discrimination mechanisms to be introduced. Continuous monitoring frameworks monitor performance measures longitudinally, reporting degradation in fairness properties that initiate remediation processes [10].

Ethical audits being as ubiquitous as financial audits signifies the fullness of AI governance from a new practice to a professionalized profession. Specialist audit firms and trained independent experts will perform thorough assessments of algorithmic systems, reviewing technical characteristics, governance procedures, and outcome trends. Audit scope will include model validation, fairness testing, explainability evaluation, data quality audit, governance effectiveness assessment, and regulatory compliance check. Audit reports will issue formal opinions on whether institutions have sufficient controls over risks of algorithmic bias, as financial audit opinions have issued about the effectiveness of internal controls. Algorithmic auditing will professionalize through developing standardized methods, auditor certification programs, and ethical standards for the practice of auditing.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The competitive edge will move to institutions that not only implement AI but do so responsibly, as regulators and consumers increasingly prefer ethical conduct to technical wizardry. Financial institutions that exhibit fairness leadership will lure socially responsible customers, especially younger consumers who value corporate social responsibility. Ethical AI practices will minimize regulatory attention and enforcement risk, reducing compliance expenses and reputational risk. Organizations with solid fairness practices will draw top talent from ethical technology and data science professionals who want to work for responsible innovators.

Market forces will increasingly benefit those who adopt AI responsibly across various channels. Environmental, social, and governance considerations in investment portfolios by institutional investors will promote financial services companies that exhibit ethical AI conduct. Consumer groups will release ratings of institutional fairness scores, which will drive customer acquisition and retention [10]. Competitive differentiation will arise not only from the deployment of sophisticated AI capabilities but from the evidence of responsible deployment by clear governance, rigorous testing for fairness, and accountable decision-making mechanisms. Organizations that invest early in ethical AI infrastructure will create reputational strengths and operational competencies that are hard for others to follow, building sustainable competitive moats on trustworthiness instead of technical superiority alone.

Development Area	Emerging Practices	Expected Outcomes
Fairness Auditing	Standardized testing protocols, pre-deployment validation, continuous monitoring requirements, and disparate impact measurement	Mandatory compliance obligations, regulatory supervision enforcement, and voluntary commitment transformation
Governance Architecture	Hybrid human-AI structures, borderline case review, confidence threshold implementation, expert panel examination	Workflow-embedded oversight, critical decision point positioning, efficiency-judgment balance
Impact Assessments	Comprehensive evaluations across fairness dimensions, stakeholder consultation, quantitative disparity analysis, and mitigation planning	Environmental review equivalency, standardized methodologies, comparative institutional evaluation
Transparency Reporting	Detailed architecture disclosures, training data characterizations, validation methodology descriptions, demographic performance metrics	External scrutiny enablement, competitive fairness pressure, and accountability mechanism creation

Table 4: Future Evolution of AI Credit Governance and Oversight [9, 10]

Conclusion

The path of AI-based credit decisioning is a turning point for financial services, in which the capabilities of technology have gotten ahead of ethical frameworks and regulatory structures architected for previous periods. Banking institutions have come to a juncture where choices today will set the course for whether algorithmic underwriting will be an engine of economic inclusion or a tool for encoding discrimination on unprecedented scales. The inevitability of the adoption of AI drives the inquiry not if but how to use these technologies in ways that promote responsible predictive power for the common good and not perpetuate biases in training data to enshrine historical inequities. Ethical design needs to move to a core position from being an add-on, and fairness considerations need to be included throughout the whole life cycle, from data gathering to model building, deployment, monitoring, and ongoing improvement. The four-

pillar structure offers prescriptive recommendations for institutions to manage tensions between responsibility and innovation, presenting specific practices in inclusive data sourcing, explainable model development, equitable governance deployment, and effective human oversight. Competitive success will build more and more on institutions that evidence responsible AI practices, as regulators will require transparency, consumers will anticipate accountability, and society will examine results across demographic axes. Market pressures will incentivize integrity in addition to technical expertise, with institutional investors, consumer protection groups, and regulators establishing multiple pressure points in favor of ethical behavior. Professionalization of algorithmic auditing, fairness metric standardization, and required impact assessments will make AI governance a mature practice similar in rigor to financial auditing. Institutions that invest actively in ethical infrastructure will create reputational assets and operating capabilities that are difficult for competitors to replicate, building sustainable differentiation upon legitimacy and social license to operate. The imperative is urgent, and action must be taken today because responses delayed only compound risk, while early leadership puts institutions ahead in shifting regulatory environments and altering consumer demands. The future of credit will not be characterized by predictive power but essentially by the capacity to produce fair, transparent, and responsible outcomes benefiting all classes of society in a balanced manner, evolving financial access from a driver of continued disadvantage into a true avenue for economic opportunity and intergenerational prosperity building.

References

[1] Robert Bartlett, et al., "CONSUMER-LENDING DISCRIMINATION IN THE FINTECH ERA," NBER WORKING PAPER SERIES, 2019. [Online]. Available: https://www.nber.org/system/files/working_papers/w25943/w25943.pdf

[2] Debidutta Pattnaik, et al., "Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review," Heliyon, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844023107006

[3] Mohamed Ali Mestikou, et al., "Artificial intelligence and machine learning in financial services: Market developments and financial stability implications," ResearchGate. [Online]. Available: https://www.researchgate.net/profile/Yassine-Hachaichi/publication/369978046_Artificial_intelligence_and_machine_learning_in_financial_services_Market_developments_and_financial_stability_implications/links/6437d64e4e83cd0e2facd021/Artificial-intelligence-and-machine-learning-in-financial-services-Market-developments-and-financial-stability-implications.pdf

[4] Ashesh Rambachan and Jonathan Roth, "A More Credible Approach to Parallel Trends," Git Hub, 2022. [Online]. Available: https://asheshrambachan.github.io/assets/files/hpt-draft.pdf

[5] Matúš Mesarčík, et al., "Stance on The Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence – Artificial Intelligence Act," ResearchGate, 2022. [Online]. Available: https://www.researchgate.net/publication/359116983_Stance_on_The_Proposal_for_a_Regulation_Laying_Down_Harmonised_Rules_on_Artificial_Intelligence_-_Artificial_Intelligence_Act

[6] Kristian Lum and William Isaac, "To predict and serve?" Wiley Online, 2016. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2016.00960.x?shared_access_token=IDAY641RsV7xsRV6B1hl04ta6bR2k8jH0KrdpFOxC677U02ZpRJGfPu343uxAHkqNhmTYCA7Luw-6h3KOpptkCw153jVJOMXzw25e_4e82JLu75RVoBQza76WyvWFk7BJN2b0K0j-zwfXRgYNqUDbybnAVsHSC6KhR9WXrKYTMY%3D

Research Article

[7] Ayse Demir, et al., "Fintech, financial inclusion and income inequality: a quantile regression approach," The European Journal of Finance, 2020. [Online]. Available: https://www.tandfonline.com/doi/pdf/10.1080/1351847X.2020.1772335
[8] Charlotte Freitag, et al., "The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations," Patterns, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389921001884
[9] Andrew D. Selbst, et al., "Fairness and Abstraction in Sociotechnical Systems," ACM Digital Library, 2019. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3287560.3287598
[10]Luca Oneto and Silvia Chiappa, et al., "Fairness in Machine Learning," arXiv, 2020. [Online]. Available: https://arxiv.org/pdf/2012.15816