Algorithmic Opacity, Proxy Discrimination and Financial Citizenship in AI-Driven Credit Markets

DOI: https://doi.org/10.58766/rpgbcb.v20i1.1277

Pedro Lima Marcheri*

Received: 20/4/2026
Approved: 7/8/2026

Introduction. 1 Artificial intelligence and the transformation of credit decision-making. 2 Proxy discrimination, algorithmic opacity and the legal governance of AI credit. 3 Algorithmic opacity as a crisis of financial governance: evidence and institutional analysis. 3.1 The Apple Card case: a catalyst for public and regulatory scrutiny. 3.2 Supervisory capacity and the governance gap: central bank mandates and the legitimacy of AI-driven finance. Conclusion.

Abstract

This article investigates the intersection of artificial intelligence, consumer credit, and financial governance, arguing that algorithmic opacity in AI-driven credit scoring does not eliminate discrimination but transforms it into an epistemic and governance crisis. Methodologically, the research adopts a qualitative, bibliographic-documentary approach grounded in doctrinal legal analysis and comparative law. Data were gathered from legislation, sectoral regulation, peer-reviewed scholarship, and institutional reports produced by the Banco Central do Brasil (BCB), *the Bank for International Settlements, the World Bank, and the U.S. Government Accountability Office. The Apple Card controversy is examined as an instrumental case study based on the official report of the New York State Department of Financial Services (2021). The comparative analysis encompasses the regulatory frameworks of Brazil, the European Union, and the United States, allowing for the identification of structural convergences and divergences in institutional responses to algorithmic opacity. The study demonstrates that historical biases are statistically reconstructed through proxy variables while the inherent opacity of machine learning models conceals these processes under a veneer of data-driven objectivity. The result is a structural deficit in evidence, contestability, and supervisory capacity that undermines the core pillars of financial citizenship: protection, participation, and trust. The original contribution lies in reconceptualizing algorithmic credit discrimination as a problem produced by opacity rather than solely by intentional disparate treatment, revealing that the absence of evidence of bias cannot be equated with evidence of fairness when the system under scrutiny is designed in ways that resist independent verification.*

Keywords: *artificial intelligence; credit scoring; algorithmic opacity; proxy discrimination; financial citizenship.*

* *Holds a Doctorate in Law from Instituição Toledo de Ensino (ITE) in Bauru. Holds a Master's degree in Law from Centro Universitário Eurípedes de Marília (UNIVEM). Completed a specialization in Criminal Law and Criminal Procedural Law at Faculdade de Direito Damásio de Jesus. Holds a law degree from Instituição Toledo de Ensino (ITE). Criminal defense attorney with experience in the legal field – specifically in Criminal Law – focusing primarily on the topics of Nazism and racial discrimination. Author of national and international publications. Permanent reviewer for the Revista da Faculdade de Direito de Franca (UNESP). Undergraduate professor of Criminal Law at Centro Universitário Católico Salesiano "Auxilium" (Unisalesiano) in Lins. http://orcid.org/0000-0003-0301-2819*

Opacidade Algorítmica, Discriminação por *Proxies* e Cidadania Financeira em Mercados de Crédito Baseados em Inteligência Artificial

Resumo

Este artigo investiga a interseção entre inteligência artificial, crédito ao consumidor e governança financeira, sustentando que a opacidade algorítmica em sistemas de pontuação de crédito baseados em IA não elimina a discriminação, mas a transforma em uma crise epistêmica e de governança. Metodologicamente, a pesquisa adota abordagem qualitativa de natureza bibliográfico-documental, com delineamento jurídico-dogmático e método comparativo. Os dados foram coletados a partir de legislação, regulação setorial, literatura acadêmica especializada e relatórios institucionais de organismos como o Banco Central do Brasil (BC), o Banco de Compensações Internacionais, o Banco Mundial e o Government Accountability Office dos Estados Unidos. O caso Apple Card foi utilizado como estudo de caso analítico-instrumental, com base no relatório oficial do New York State Department of Financial Services (2021). A análise comparada abrange os marcos regulatórios do Brasil, da União Europeia e dos Estados Unidos, permitindo identificar convergências e divergências nas respostas institucionais à opacidade algorítmica. O estudo demonstra que vieses históricos são estatisticamente reconstruídos por meio de variáveis substitutas, enquanto a opacidade inerente aos modelos de aprendizado de máquina oculta esses processos sob uma aparência de objetividade orientada por dados. O resultado é um déficit estrutural de evidência, de contestabilidade e de capacidade de supervisão que compromete os pilares centrais da cidadania financeira: proteção, participação e confiança. A contribuição original consiste em reconceituar a discriminação algorítmica no crédito como um problema produzido pela opacidade, e não apenas por tratamento discriminatório intencional, revelando que a ausência de evidência de viés não pode ser equiparada à evidência de equidade quando o sistema analisado é concebido de forma a resistir à verificação independente.

Palavras-chave: inteligência artificial; pontuação de crédito; opacidade algorítmica; discriminação por *proxies*; cidadania financeira.

Introduction

The deployment of artificial intelligence in consumer credit markets presents one of the most consequential governance challenges of the contemporary financial era. Across jurisdictions, financial institutions are rapidly migrating from traditional, human-driven underwriting processes to sophisticated machine learning systems that promise enhanced predictive accuracy, operational efficiency, and the expansion of credit access to historically underserved populations. This technological transformation, however, does not eradicate discrimination. It reconstitutes discrimination, embedding historical and societal biases within complex statistical architectures that render such biases invisible to consumers, regulators, and often to the institutions themselves.

This article argues that the central legal problem in AI-driven credit markets is not limited to discriminatory intent or even to measurable disparate impact. The deeper issue is that algorithmic opacity makes it difficult to see how decisions are produced, challenged, and supervised. When a credit model cannot be meaningfully understood

by consumers, regulators, or even the institutions that deploy it, the result is not merely a technical limitation but a deficit of accountability. In that setting, the absence of evidence of bias cannot be taken as evidence of fairness.

This research adopts a qualitative design of bibliographic-documentary nature. The primary methodological approach is doctrinal legal analysis, which examines the normative content, internal coherence, and operative scope of applicable legal rules, principles, and regulatory frameworks. This method is complemented by comparative law analysis, applied to the credit governance regimes of Brazil, the European Union, and the United States, with the purpose of identifying structural convergences and divergences in their respective institutional responses to algorithmic opacity and proxy discrimination.

Data collection drew upon four categories of sources: primary legal sources, including statutes, resolutions, and regulatory guidance; peer-reviewed academic literature in law, finance, and computer science; institutional and supervisory reports published by the Banco Central do Brasil, *the Bank for International Settlements, the World Bank, the U.S. Government Accountability Office, and other regulatory bodies; and official investigative documentation, specifically the 2021 report of the New York State Department of Financial Services on the Apple Card investigation.*

The Apple Card case is employed as an instrumental case study; not as a representative sample, but as an analytically illuminating instance that brings theoretical propositions into contact with documented regulatory and institutional evidence. The selection criterion was the case's unique capacity to operationalize the abstract concepts of algorithmic opacity and epistemic governance failure within a formally investigated and publicly documented context.

The analytical framework integrates legal-normative reasoning with institutional and empirical analysis, assessing regulatory adequacy against the standard of meaningful explainability, contestability, and supervisory capacity. This multi-layered approach is intended to ensure that the conclusions are grounded in both doctrinal validity and empirical plausibility, and that the study is replicable by researchers with access to the same publicly available sources.

The analysis combines doctrinal legal reasoning with a comparative examination of Brazil, the European Union, and the United States. It also draws on the Apple Card investigation as a practical illustration of how opaque systems can generate public controversy even when regulators do not find unlawful discrimination. Institutional reports from bodies such as the Bank for International Settlements, the World Bank, the U.S. Government Accountability Office, and the Banco Central do Brasil *are used to assess the supervisory implications of these systems.*

The article is organized in three parts. The first examines how AI, datafication, and regulatory innovation are reshaping credit markets, with particular attention to the Brazilian experience and the idea of financial citizenship. The second focuses on proxy discrimination and algorithmic opacity, comparing the legal frameworks of Brazil, the European Union, and the United States, and considering the limits of explainable AI as a governance tool. The third uses the Apple Card case and institutional evidence to show how opacity can weaken contestability, public trust, and supervisory capacity. The conclusion reflects on what this means for the future of credit governance in AI-driven markets.

1 Artificial intelligence and the transformation of credit decision-making

The Brazilian credit market is undergoing a deep structural change, driven by the convergence of regulatory innovation and the datafication of financial intermediation. Historically characterized by high intermediation costs, stark banking concentration, and persistent information asymmetry, this landscape is being redefined by the widespread adoption of artificial intelligence and machine learning in credit decision-making. This transformation is powered by an expanding universe of traditional and, increasingly, alternative data, shifting the very foundation upon which creditworthiness is assessed and allocated.

CC BY NC SA

The basis of modern credit has always been risk assessment, but the methodology of that assessment has evolved dramatically. Traditionally, credit scoring operated as a retrospective mechanism grounded in past repayment behavior and historical financial records, which systematically excluded individuals lacking formal credit histories and produced large populations of so-called "thin-file" consumers. In emerging economies such as Brazil, characterized by high levels of informality, this structural limitation has historically represented a persistent barrier to financial inclusion and access to credit markets (Ornelas; Oliveira; Schechtman, 2025, p. 3–5). The contemporary transformation is both conceptual and technological: credit evaluation increasingly shifts from documenting past financial conduct toward predicting future repayment capacity through large-scale data processing. This transformation is driven by the datafication of finance, in which lenders incorporate alternative data sources, including transactional information, digital behavior, mobile usage patterns, and other non-traditional indicators, to infer creditworthiness (Berg et al., 2020, p. 2846–2847). Machine learning systems analyze these heterogeneous datasets using complex and non-linear predictive structures that surpass the explanatory and predictive limits of traditional statistical scoring models, enabling more granular and behaviorally informed risk assessments (Das; Stanton; Wallace, 2022, p. 565–568).

A pivotal catalyst in Brazil was Complementary Law No. 166/2019, which consolidated the adoption of an opt-out model for the positive credit bureau system (cadastro positivo). The reform required financial institutions and other data furnishers to automatically share consumer payment histories with credit bureaus, expanding credit information beyond records of default to include positive repayment behavior. The policy objective was to reduce informational asymmetries in lending markets and improve risk assessment through broader data availability, thereby facilitating credit expansion and lowering borrowing costs (Ornelas; Oliveira; Schechtman, 2025, p. 3–7). By incorporating positive information sharing, the cadastro positivo represents a regulatory strategy designed to mitigate the structural limits of retrospective credit evaluation and to build a richer informational environment capable of supporting more accurate predictive assessment models. In this sense, the reform created the informational infrastructure necessary for data-driven and AI-based credit scoring systems to operate under conditions of greater market coverage and potential financial inclusion (Ornelas; Oliveira; Schechtman, 2025, p. 8–10).

This data-rich environment has paved the way for customized, purpose-built scoring models that move away from a one-size-fits-all approach toward more inclusive and flexible frameworks, empowering consumers with greater control over their financial identity. Yet the very richness and granularity of the data that fuels these models also amplifies the risks of proxy discrimination.

The technological and data-driven transformation of credit scoring in Brazil has been enabled by the emergence of a digitally native financial infrastructure shaped by coordinated regulatory innovation. This ecosystem, frequently described as the Brazil Stack, reflects the integration of public regulatory initiatives with private technological development, redefining how financial services are produced, distributed, and consumed within the national financial system (McKinsey & Company, [s. d.], p. 6–9). Its central components include the instant payment system Pix, the implementation of Open Finance, and expanding digital identification mechanisms, which together have created institutional and technological conditions favorable to the expansion of financial technology firms and to the consolidation of artificial intelligence as an operational infrastructure within financial markets (FGV Rio Law, [s. d.], p. 2–4).

The rapid expansion of financial technology adoption in Brazil reflects broader structural changes in financial intermediation. Increased digitalization, regulatory openness, and competitive pressures have encouraged financial institutions to integrate advanced data analytics and automated decision-making systems into fraud detection, cybersecurity, credit risk management, and customer interaction processes (World Bank, 2024, p. 28–31). Over the last decade, the growth of fintech firms has intensified competition by introducing mobile-first service models, lower operational costs, and data-driven products, prompting incumbent banks to accelerate technological investment and digital transformation strategies across the sector (International Monetary Fund, 2025, p. 4–6). This competitive dynamic has contributed to greater efficiency in credit allocation and declining intermediation costs, reinforcing

the role of technological innovation as a driver of structural change in Brazil's financial markets (World Bank, 2024, p. 28–31).

A central pillar of this new infrastructure is Pix, the instant payment system launched by the Banco Central do Brasil *in 2020. The system substantially expanded access to digital payments by enabling real-time transfers at low cost, increasing financial participation and formalizing transactional activity previously concentrated in cash-based exchanges. The large-scale adoption of instant payments has generated extensive transactional datasets, strengthening informational availability for financial institutions and supporting the development of more data-intensive credit assessment practices (*Banco Central do Brasil, *2021, p. 94–95; FGV Rio Law, [s. d.], p. 16).*

Although technological innovation supplied the operational tools for transformation, the decisive enabling factor was a deliberate regulatory strategy centered on data governance and competition policy. The implementation of Open Finance by the Banco Central do Brasil *represents a structural reform aimed at reshaping the architecture of the national financial system by promoting interoperability, transparency, and user control over financial data (FGV Rio Law, [s. d.], p. 3–4). The initiative has been widely regarded as an advanced regulatory model due to its comprehensive scope, rapid implementation, and emphasis on consumer empowerment within data-sharing ecosystems (World Bank, 2024, p. 32–34).*

*The primary objective of Open Finance is to increase efficiency, innovation, and competitive neutrality by allowing consumers to authorize the secure transfer of their financial information between institutions through standardized application programming interfaces. By enabling portability of registration data, transactional histories, and credit information, the framework reduces historical informational asymmetries that favored incumbent institutions holding proprietary datasets and limited market entry by new competitors (*Banco Central do Brasil, *[s.d].; World Bank, 2024, p. 33).*

The implementation followed a phased regulatory strategy designed to progressively expand data sharing and functional interoperability. Initial stages required institutions to disclose standardized information about products and services, followed by consumer-authorized sharing of transactional data and the introduction of payment initiation services. Subsequent phases broadened the framework beyond traditional banking activities to encompass insurance, investment, and pension data, consolidating the transition from Open Banking to a comprehensive Open Finance ecosystem structured around integrated financial services and advanced data governance mechanisms (FGV Rio Law, [s. d.], p. 3–4; p. 16).

*By democratizing access to financial information, Open Finance reduces informational barriers that historically protected incumbent institutions and facilitates market entry by fintech firms capable of competing through data-driven services and digital distribution models (FGV Rio Law, [s. d.], p. 3–4; World Bank, 2024, p. 33). The framework allows authorized institutions to access standardized consumer financial data across multiple providers, enabling more comprehensive risk analysis and individualized financial products grounded in broader transactional visibility rather than isolated credit bureau records (*Banco Central do Brasil, *s.d.). This expanded data environment supports increasingly personalized credit offerings and adaptive financial services aligned with observed income flows, consumption behavior, and existing financial obligations.*

Open Finance functions as an integrating layer within Brazil's broader digital financial infrastructure by allowing data generated through payment systems such as Pix to circulate securely under consumer consent and standardized governance protocols. While this reduces informational asymmetries between financial institutions, it simultaneously introduces a new imbalance between technologically sophisticated lenders equipped with advanced analytical capabilities and individual consumers who often lack equivalent interpretive or bargaining capacity, raising governance and consumer protection concerns associated with algorithmic decision-making (World Bank, 2024, p. 34; FGV Rio Law, [s. d.], p. 16).

The technological and regulatory transformations observed in Brazil's credit market are linked to a broader institutional objective promoted by the Banco Central do Brasil, *namely the advancement of Financial Citizenship.*

CC BY NC SA

*The concept is defined as an integrated policy framework encompassing financial inclusion, financial education, consumer protection, and citizen participation within the financial system, reflecting a shift from mere access to financial services toward sustainable and informed use of those services (*Banco Central do Brasil, *2021, p. 7–9).*

*This policy agenda is monitored through the Relatório de Cidadania Financeira, which replaced earlier inclusion-focused monitoring frameworks in order to incorporate broader dimensions of well-being, participation, and trust in financial markets (*Banco Central do Brasil, *2021, p. 4–5). International comparative evidence, including data compiled by the World Bank, indicates substantial expansion of account ownership in Brazil during recent years, reflecting the combined effects of digital payment infrastructure, social transfer programs delivered through electronic accounts, and regulatory modernization of the financial ecosystem (World Bank, 2024, p. 12–14). Administrative evidence from the Banco Central similarly demonstrates the rapid expansion of financial relationships mediated by payment institutions and digital platforms, reinforcing the role of technological infrastructure in accelerating financial inclusion (*Banco Central do Brasil, *2021, p. 94–100).*

*Despite these advances, structural challenges remain evident. Reports on financial citizenship continue to identify perceived account costs, income instability, and limited financial capability as persistent barriers to full participation in formal financial markets, indicating that expanded access alone does not guarantee effective or sustainable financial integration (*Banco Central do Brasil, *2021, p. 98–100). Concerns regarding financial resilience and household capacity to manage recurring obligations further highlight the importance of financial education and consumer protection policies as complementary components of inclusion strategies (*Banco Central do Brasil, *2021, p. 94–95).*

Within this framework, financial citizenship provides the normative foundation for evaluating governance challenges associated with algorithmic credit allocation. As artificial intelligence systems increasingly mediate access to financial opportunities, questions of transparency, accountability, and procedural fairness become central to ensuring that individuals can meaningfully participate in and trust financial institutions operating through automated decision-making processes (Aggarwal, 2021, p. 53–54).

The accelerated integration of artificial intelligence into financial decision-making has therefore prompted the development of new regulatory approaches in Brazil. Ongoing legislative initiatives aimed at establishing a national legal framework for artificial intelligence adopt a risk-based regulatory logic comparable to international models, classifying credit scoring and other socially consequential applications as high-risk systems subject to heightened governance requirements, including transparency obligations, explainability standards, and human oversight mechanisms. Such proposals classify credit scoring among the applications subject to reinforced governance obligations, reflecting a broader international convergence toward risk-based regulation of algorithmic systems (Aggarwal, 2021, p. 53-54; Prince; Schwarcz, 2020, p. 1273-1275).

A central regulatory concern relates to the opacity inherent in complex machine learning systems. Contemporary legal scholarship emphasizes that meaningful accountability requires institutions to provide understandable explanations for automated decisions and to maintain internal governance structures capable of auditing algorithmic behavior and mitigating discriminatory outcomes (Das; Stanton; Wallace, 2022, p. 582–585; Prince; Schwarcz, 2020, p. 1273–1275).

Brazilian consumer protection doctrine traditionally attributes objective liability to financial institutions for risks arising from their economic activities, a principle that extends to harms generated by automated systems when algorithmic decisions produce unfair or discriminatory outcomes. Applied to automated decision systems, this principle reinforces that institutions cannot displace responsibility onto the opacity of the technological architectures they deploy.

Regulatory authorities such as the Banco Central do Brasil *and the Autoridade Nacional de Proteção de Dados are expected to occupy central roles in supervising algorithmic governance within the financial sector. Existing prudential supervision already requires documentation, validation, and continuous monitoring of credit risk models,*

reflecting a broader movement toward supervisory oversight of automated decision systems (Banco Central do Brasil, *2020, p. 1-5).*

This emerging hybrid governance model combines preventive regulation, supervisory oversight, and institutional accountability mechanisms aimed at strengthening trust in data-driven financial ecosystems. Empirical and theoretical research indicates that demonstrable algorithmic fairness increasingly functions not only as a legal compliance requirement but also as a strategic factor influencing consumer confidence and competitive positioning within digitally mediated financial markets (Hurlin; Pérignon; Saurin, 2026, p. 420–425).

2 *Proxy discrimination, algorithmic opacity and the legal governance of AI credit*

The integration of artificial intelligence into credit decision-making introduces two interrelated challenges central to contemporary financial governance: algorithmic opacity and proxy discrimination. The effectiveness of modern credit models, including neural networks and ensemble methods, derives from their ability to process large volumes of heterogeneous data and identify complex non-linear relationships that extend beyond the capabilities of traditional statistical approaches (Ariza-Garzón et al., 2020, p. 64873–64876). This technical advantage, however, is accompanied by a well-documented trade-off between predictive performance and interpretability, in which more accurate models tend to be less transparent and harder to explain (Aggarwal, 2021, p. 48).

This trade-off underlies the so-called black box problem, in which the relationship between inputs and outputs is observable but the internal decision-making logic of the model remains opaque. High-performing machine learning systems often rely on intricate feature interactions and layered computations that are not readily interpretable, even by their developers, limiting the ability of external auditors or affected individuals to understand how specific decisions are reached (Ariza-Garzón et al., 2020, p. 64884–64890). This opacity is not merely a technical limitation but a structural obstacle to accountability, particularly in regulatory environments that require institutions to provide clear and specific justifications for adverse credit decisions (Aggarwal, 2021, p. 53–54).

Within this opaque environment, proxy discrimination emerges as a significant risk. Even when protected attributes such as race or gender are excluded from model inputs, algorithms may rely on correlated variables that function as indirect proxies for these characteristics, reproducing patterns of structural inequality embedded in historical data (Prince; Schwarcz, 2020, p. 1257–1260). Empirical research in credit markets demonstrates that seemingly neutral variables can encode socioeconomic and demographic information, allowing models to replicate discriminatory outcomes without explicit intent (Avery; Brevoort; Canner, 2010, p. 3–4). This dynamic obscures discriminatory effects behind an appearance of technical objectivity, making them difficult to detect and challenge (Prince; Schwarcz, 2020, p. 1273–1275).

The opacity of algorithmic systems significantly constrains the ability of individuals and regulators to identify and prove the use of such proxies, creating an epistemic barrier to enforcement and accountability. Rather than eliminating discrimination, automated systems risk embedding it in forms that are less visible and more difficult to contest. This challenge has prompted different regulatory responses across jurisdictions.

In the United States, the governance of algorithmic credit systems relies on the application of existing legal frameworks rather than comprehensive AI-specific legislation. Anti-discrimination principles embedded in financial regulation impose obligations on lenders to justify adverse decisions and ensure that credit practices do not produce unlawful disparate impacts, requirements that are difficult to reconcile with opaque algorithmic models (Aggarwal, 2021, p. 53–54). Similarly, regulatory expectations regarding the disclosure of key factors influencing credit decisions reinforce the need for interpretability in automated systems, highlighting tensions between legal transparency requirements and technical complexity (Das; Stanton; Wallace, 2022, p. 582–585).

CC BY NC SA

The European Union has adopted a more comprehensive approach grounded in fundamental rights and ex ante regulation. Data protection law establishes safeguards against fully automated decision-making and requires that individuals receive meaningful information about the logic underlying such processes, particularly when decisions produce significant legal or economic effects (Aggarwal, 2021, p. 53–54). More recent regulatory developments extend this framework by classifying credit scoring systems as high-risk applications, thereby imposing obligations related to risk management, data governance, transparency, and human oversight. These developments are situated within a broader European framework for the governance of high-risk AI systems in credit markets (Kalus, 2025).

*Brazil's regulatory approach combines data protection principles with sector-specific financial supervision, creating a layered governance structure. The Lei Geral de Proteção de Dados establishes rights to review automated decisions and embeds principles of transparency and non-discrimination in data processing activities, aligning with international standards (Prince; Schwarcz, 2020, p. 1273–1275). At the same time, financial regulation imposes requirements for documentation, validation, and monitoring of credit risk models, reflecting a supervisory concern with the risks associated with opaque and complex decision systems. These requirements find explicit expression in prudential regulations applicable to credit fintechs and to risk-based supervision of financial institutions (*Banco Central do Brasil, *2020, p. 1-5).*

The principle of objective liability in consumer protection law further reinforces accountability by attributing responsibility to financial institutions for harms arising from their services, including those generated by automated systems.

In response to the challenges posed by opacity, the development of explainable artificial intelligence has sought to reconcile predictive performance with interpretability. Techniques such as SHAP and LIME provide post hoc explanations by estimating the contribution of input variables to individual predictions or approximating complex models with simpler, locally interpretable representations (Gramegna; Giudici, 2021; Ariza-Garzón et al., 2020, p. 64884–64890). These methods can support regulatory compliance by enabling more transparent communication of decision factors and facilitating auditing processes (Das; Stanton; Wallace, 2022, p. 589–593).

Despite their practical utility, such techniques have important limitations. Post hoc explanations do not fully reveal the internal logic of complex models but instead provide approximations that may diverge from the actual decision process. This limitation raises questions about whether current explainability tools satisfy legal standards requiring meaningful transparency, particularly in high-stakes contexts such as credit allocation (Das; Stanton; Wallace, 2022, p. 582–585).

Without effective interpretability mechanisms, legal rights related to explanation, contestation, and non-discrimination risk becoming largely formal rather than substantive. In this context, opacity is not simply a technical characteristic but a structural feature that shapes the distribution of power, knowledge, and accountability in AI-driven credit markets.

3 Algorithmic opacity as a crisis of financial governance: evidence and institutional analysis

3.1 The Apple Card case: a catalyst for public and regulatory scrutiny

The theoretical challenges of algorithmic opacity and proxy discrimination found concrete empirical expression in the Apple Card controversy of 2019, which brought algorithmic bias in credit decision-making into global public debate. The Apple Card, issued by Apple Inc. and underwritten by Goldman Sachs Bank USA, became the subject of allegations of sex-based discrimination following widely publicized consumer complaints (New York State Department of Financial Services, 2021, p. 3–5).

The controversy began when a technology entrepreneur reported receiving a credit limit substantially higher than that granted to his wife despite shared finances and joint tax filings, attributing the discrepancy to potential gender bias embedded in the underwriting algorithm. The issue gained further attention when Apple co-founder Steve Wozniak described a similar experience, amplifying concerns about opaque automated credit systems and triggering a formal investigation by the New York State Department of Financial Services, the primary regulator overseeing Goldman Sachs Bank (New York State Department of Financial Services, 2021, p. 5–7).

The regulatory investigation illustrates the complexity of governing algorithmic decision-making. After reviewing extensive documentation, analyzing underwriting data from hundreds of thousands of applicants, and interviewing complainants, the regulator concluded that there was no evidence of unlawful disparate treatment or disparate impact discrimination against women. Statistical analysis demonstrated that applicants with comparable credit characteristics received similar outcomes regardless of gender, and the institution was able to justify decisions using conventional risk variables such as credit score, indebtedness, income, and repayment history (New York State Department of Financial Services, 2021, p. 10–18).

Despite the absence of legal discrimination, the investigation identified significant shortcomings in transparency and consumer communication. Although the bank could technically explain its models to regulators, it failed to provide understandable explanations to customers. Consumers encountered customer service representatives unable to clarify why credit limits differed, reinforcing perceptions of an unaccountable black box system. The regulator observed that clearer explanatory policies could have mitigated confusion and mistrust, particularly regarding credit limits, even where formal legal explanations were not strictly required (New York State Department of Financial Services, 2021, p. 19–22).

The investigation also clarified a widespread misconception among complainants that spouses sharing finances must receive identical credit terms. The regulator found that spouses frequently maintain distinct individual credit histories and risk profiles, which legally justify divergent lending outcomes despite shared assets or joint tax filings (New York State Department of Financial Services, 2021, p. 23–24).

The analytical importance of the Apple Card case lies less in its legal conclusion than in what it reveals about governance under algorithmic opacity. The controversy demonstrates a shift from disputes centered on provable discrimination toward what may be characterized as an epistemic crisis: even in the absence of regulatory violations, opaque systems generate informational asymmetries that erode public trust. The lack of transparency created an evidentiary vacuum in which regulatory findings failed to resolve public suspicion because the underlying decision process remained inaccessible (New York State Department of Financial Services, 2021, p. 24–26).

This phenomenon reflects a structural limitation of opaque algorithmic systems with direct implications for legal doctrine and supervisory practice. When the internal logic of a credit model cannot be meaningfully scrutinized by consumers or clearly articulated by institutions, contestability, a foundational component of procedural fairness in consumer finance, becomes weakened. Legal regimes across jurisdictions formally guarantee rights to explanation and review of automated decisions, yet these guarantees lose effectiveness when explanations cannot convey the causal reasoning behind algorithmic outcomes (Prince; Schwarcz, 2020, p. 1273–1275; Aggarwal, 2021, p. 53–54).

*Under such conditions, a legally protected right risks becoming merely formal rather than substantive. Within Brazil's financial citizenship framework, this erosion of contestability directly undermines both consumer protection and participatory inclusion, since individuals cannot effectively exercise rights within systems whose functioning they cannot understand or challenge (*Banco Central do Brasil, *2021, p. 9–12).*

CC BY NC SA

3.2 Supervisory capacity and the governance gap: central bank mandates and the legitimacy of AI-driven finance

The challenge for supervisors is acute and multifaceted. Regulators face a formidable barrier to oversight when trying to audit systems whose internal logic is opaque. In high-stakes decisions, the literature argues for inherently interpretable models rather than post hoc explanations, and it shows that better statistical technology in credit markets can create more dispersed predictions and, consequently, winners and losers even when the model is simply more accurate (Rudin, 2019, p. 206-215; Fuster et al., 2022, p. 8-9). Data mining can also rely on proxy variables that systematically disadvantage protected groups and make discrimination difficult to detect (Barocas; Selbst, 2016, p. 671-673, 691-692).

The 2025 GAO report found that federal financial regulators are still overseeing AI primarily through existing laws, guidance, and risk-based examinations, while financial institutions remain cautious about using less explainable models in high-reliability, customer-facing decisions; it also identified NCUA's third-party oversight gap (U.S. Government Accountability Office, 2025, p. 1-2, 12, 32-33). The 2026 interagency model-risk guidance from the OCC, Federal Reserve, and FDIC is explicitly risk-based and tailored, but generative AI and agentic AI are outside its scope, and the guidance emphasizes validation, monitoring, governance, and third-party products rather than prescriptive rules (Office of the Comptroller of the Currency, 2026, p. 1-3). The BIS notes that dependency on a few large technology providers can complicate oversight and that many authorities still prefer a technology-neutral approach, while the World Bank argues that AI may be a game-changer for risk-based supervision, especially where human resources are limited (Bank for International Settlements, 2024, p. 6, 29-30; World Bank, 2025, p. 7, 13).

The erosion of contestability and supervisory capacity inevitably degrades consumer trust, which is the bedrock of a stable financial system. In the Apple Card investigation, the New York State Department of Financial Services found no fair lending violations, but it also concluded that customer-service deficiencies and a perceived lack of transparency undermined consumer trust in fair credit decisions (New York State Department of Financial Services, 2021, p. 1-2). In the context of financial citizenship, the Banco Central do Brasil *defines the concept as the exercise of rights and duties that depends on a structured environment of inclusion, education, consumer protection, and participation, and its 2021 report recasts the agenda around inclusion, competitiveness, transparency, education, and sustainability in a more digital financial system (*Banco Central do Brasil, *2021, p. 6-7;* Banco Central do Brasil, *2021, p. 4-5).*

For a central bank like the Banco Central do Brasil, *this governance crisis directly challenges core mandates. Credit-fintech resolution requires consistent, verifiable, and transparent criteria for client selection and a credit analysis model capable of providing indicators that impartially reflect debtor risk, which is precisely the kind of standard opaque models strain to satisfy (*Banco Central do Brasil, *2020, p. 1, 5). The success of Brazil's financial inclusion agenda therefore depends not only on broader access, but on the perceived integrity of the financial system's data and governance architecture, especially in a context of digital-payment expansion such as Pix (*Banco Central do Brasil, *2021, p. 123-127).*

This analysis thus reveals that the governance crisis produced by algorithmic opacity is not marginal or transitional (Rudin, 2019, p. 206-215); it is a structural feature of the current technological paradigm that requires a multi-layered response combining legal, technical, and institutional innovations.

Conclusion

The integration of artificial intelligence into credit markets has changed the nature of discrimination, shifting it from overt bias to statistical reconstruction and evidential invisibility. The central argument of this work is that algorithmic opacity produces an epistemic and governance deficit that systematically disempowers consumers and constrains regulators, a claim supported here through doctrinal legal analysis, comparative regulatory study, and empirical evidence.

Chapter 1 demonstrated that the Brazilian credit market is undergoing a profound structural transformation driven by the convergence of AI adoption, datafication, and regulatory innovation. The so-called Brazil Stack, composed of Pix, Open Finance, and an active fintech ecosystem, has expanded financial access and generated the data infrastructure on which AI-driven credit models operate. Yet this transformation is embedded within the broader normative framework of Financial Citizenship, which demands not merely inclusion but also protection, education, and meaningful participation.

Chapter 2 established the theoretical and legal foundations of the governance crisis. The twin challenges of algorithmic opacity and proxy discrimination create a structural tension with legal frameworks that demand transparency, explainability, and contestability. The comparative analysis of the United States, the European Union, and Brazil revealed distinct but converging regulatory philosophies: the U.S. relies on enforcement-driven adaptation of existing sectoral laws; the EU has adopted a comprehensive, rights-based approach with ex-ante obligations; and Brazil presents a hybrid model combining LGPD-based data protection rights with proactive central bank regulation and objective liability under the Consumer Defense Code. Explainable AI techniques offer a promising but incomplete response, providing interpretability layers that may satisfy legal requirements while raising questions about whether post-hoc approximations truly capture a model's internal logic.

Chapter 3 synthesized empirical and institutional evidence to demonstrate how opacity functions in practice as a crisis of financial governance. The Apple Card investigation revealed that even when a formal regulatory review finds no evidence of unlawful discrimination, the opacity of the system creates an information vacuum that erodes public trust and renders contestability meaningless. Institutional reports from the GAO, the BIS, and the World Bank confirmed that supervisory authorities face significant capacity constraints in auditing opaque AI systems, exacerbated by reliance on third-party technology providers and the rapid pace of technological innovation.

The original contribution of this work has been to reconceptualize algorithmic discrimination as an epistemic and governance problem. By focusing on opacity as the core mechanism, the analysis moves beyond the narrow search for intentional bias or statistically demonstrable disparate impact. It posits that opacity itself constitutes a governance failure because it makes the right to contestation meaningless and supervisory oversight untenable. The absence of evidence of discrimination, as the Apple Card investigation demonstrated, cannot be treated as evidence of fairness when the system under review is designed in a way that inherently resists independent scrutiny.

The implications of this crisis for financial citizenship are profound. The promise of AI to advance financial inclusion by scoring the previously unscorable is paradoxically undermined when the methods used erode the trust and protections that are foundational to meaningful participation in the financial system. The core pillars of financial citizenship—protection, education, participation, and inclusion—are all weakened in an environment where financial opportunities are allocated by inscrutable systems.

Ultimately, the challenge of algorithmic credit governance is not one of slowing innovation but of realigning it with the legal and ethical principles of fairness, transparency, and accountability. The analysis presented here suggests that addressing the governance crisis requires action on multiple fronts: legal frameworks that insist on meaningful explainability, not merely formal compliance; supervisory authorities equipped with the technical capacity and legal authority to audit complex systems and their third-party providers; industry commitments to building transparency

CC BY NC SA

and fairness into models by design rather than as post-hoc additions; and continued investment in Explainable AI research that bridges the gap between technical interpretability and legal standards of meaningful explanation. Without these coordinated efforts, the efficiencies gained through AI will come at the unacceptable cost of a less just, less transparent, and less legitimate financial system.

References

AGGARWAL, Nikita. The norms of algorithmic credit scoring. **The Cambridge Law Journal**, Cambridge, v. 80, n. 1, p. 42-73, 2021. Disponível em: https://www.cambridge.org/core/journals/cambridge-law-journal/article/norms-of-algorithmic-credit-scoring/23C9802EEA5EC6F6872512CB7AABC793. Acesso em: 14 mar. 2026.

ARIZA-GARZÓN, Miller Janny; ARROYO, Javier; CAPARRINI, Antonio; SEGOVIA-VARGAS, María Jesús. **Explainability of a machine learning granting scoring model in peer-to-peer lending**. IEEE Access, Piscataway, v. 8, p. 64873-64890, 2020. DOI: 10.1109/ACCESS.2020.2984412. Disponível em: https://doi.org/10.1109/ACCESS.2020.2984412. Acesso em: 20 abr. 2026.

AVERY, Robert B.; BREVOORT, Kenneth P.; CANNER, Glenn B. **Does credit scoring produce a disparate impact?** Washington, D.C.: Federal Reserve Board, 2010. (Finance and Economics Discussion Series, 2010-58). Disponível em: https://www.federalreserve.gov/pubs/feds/2010/201058/201058pap.pdf. Acesso em: 22 jan. 2025.

BANCO CENTRAL DO BRASIL. **Credit fintechs regulation – CMN Resolution No. 4,656/2020**. Brasília: Banco Central do Brasil, 2020. Disponível em: https://www.bcb.gov.br/content/config/Documents/Credit_Fintechs_Regulation_CMN_Resolution_4656_2020.pdf. Acesso em: 20 abr. 2026.

BANCO CENTRAL DO BRASIL. **Open Finance**. Brasília: Banco Central do Brasil, [s. d.]. Disponível em: https://www.bcb.gov.br/en/financialstability/open_finance. Acesso em: 20 abr. 2026.

BANCO CENTRAL DO BRASIL. **Relatório de Cidadania Financeira 2021**. Brasília: Banco Central do Brasil, 2021. Disponível em: https://www.bcb.gov.br/content/cidadaniafinanceira/documentos_cidadania/RIF/Relatorio_de_Cidadania_Financeira_2021.pdf. Acesso em: 20 abr. 2026.

BANK FOR INTERNATIONAL SETTLEMENTS. **Regulating AI in the financial sector**: recent developments and main challenges. Basel: BIS, 2024. (FSI Insights, n. 63). Disponível em: https://www.bis.org/fsi/publ/insights63.pdf. Acesso em: 20 abr. 2026.

BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. **California Law Review**, Berkeley, v. 104, n. 3, p. 671-732, 2016. Disponível em: https://www.cs.yale.edu/homes/jf/BarocasSelbst.pdf. Acesso em: 3 fev. 2026.

BERG, Tobias; BURG, Valentin; GOMBOVIĆ, Ana; PURI, Manju. On the rise of FinTechs: credit scoring using digital footprints. **The Review of Financial Studies**, Oxford, v. 33, n. 7, p. 2845-2897, 2020. Disponível em: https://academic.oup.com/rfs/article/33/7/2845/5568311. Acesso em: 19 jul. 2025.

DAS, Sanjiv; STANTON, Richard; WALLACE, Nancy. Algorithmic fairness. **Annual Review of Financial Economics**, Palo Alto, v. 14, p. 1-32, 2022. Disponível em: https://faculty.haas.berkeley.edu/stanton/pdf/ARfairness.pdf. Acesso em: 11 nov. 2025.

FGV RIO LAW. **Implementation and challenges of Open Finance in Brazil**. Rio de Janeiro: Fundação Getulio Vargas, [s. d.]. Disponível em: https://direitorio.fgv.br/sites/default/files/arquivos/direito_rio_livro_neasf_18_eng_ap5.pdf. Acesso em: 20 abr. 2026.

FUSTER, Andreas; GOLDSMITH-PINKHAM, Paul; RAMADORAI, Tarun; WALTHER, Ansgar. Predictably unequal? The effects of machine learning on credit markets. **The Journal of Finance**, Hoboken, v. 77, n. 1, p. 5-47, 2022. Disponível em: https://www.researchgate.net/publication/355710743_Predictably_Unequal_The_Effects_of_Machine_Learning_on_Credit_Markets. Acesso em: 27 ago. 2025.

GRAMEGNA, Alex; GIUDICI, Paolo. SHAP and LIME: an evaluation of discriminative power in credit risk. **Frontiers in Artificial Intelligence**, Lausanne, v. 4, art. 752558, 2021. Disponível em: https://www.frontiersin.org/articles/10.3389/frai.2021.752558/full. Acesso em: 9 jan. 2026.

HURLIN, Christophe; PÉRIGNON, Christophe; SAURIN, Sébastien. The fairness of credit scoring models. **Management Science**, Hanover, v. 72, n. 1, p. 406-425, 2026. Disponível em: https://pubsonline.informs.org/doi/10.1287/mnsc.2022.03888. Acesso em: 2 mar. 2026.

INTERNATIONAL MONETARY FUND. **Explaining strong credit growth in Brazil despite high policy rates.** Washington, D.C., 2025. Disponível em: https://www.imf.org/en/news/articles/2025/10/09/explaining-strong-credit-growth-in-brazil-despite-high-policy-rates. Acesso em: 14 ago. 2026.

KALUS. **AI systems and a risk of discrimination in creditworthiness assessments**: the role of the EU Artificial Intelligence Act in safeguarding the right to non-discrimination. Stanford: Stanford Law School, 2025. (EU Law Working Paper, n. 112). Disponível em: https://law.stanford.edu/wp-content/uploads/2025/05/EU-Law-WP-112-Kalus.pdf. Acesso em: 20 abr. 2026.

MCKINSEY & COMPANY. **Brazil's financial evolution**: unveiling the Brazil Stack. São Paulo: McKinsey & Company, [s. d.]. Disponível em: https://www.mckinsey.com/br/~/media/mckinsey/locations/south%20america/brazil/our%20insights/brazil%20stack/brazil%20stack%20report_al_vf.pdf. Acesso em: 20 abr. 2026.

NEW YORK STATE DEPARTMENT OF FINANCIAL SERVICES. **Report on Apple Card Investigation**. New York: NYDFS, 2021. Disponível em: https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf. Acesso em: 20 abr. 2026.

OFFICE OF THE COMPTROLLER OF THE CURRENCY. **Model risk management**: revised guidance (Bulletin 2026-13). Washington, D.C.: OCC, 2026. Disponível em: https://occ.gov/news-issuances/bulletins/2026/bulletin-2026-13.html. Acesso em: 20 abr. 2026.

ORNELAS, José Renato Haas; OLIVEIRA, Raquel de Freitas; SCHECHTMAN, Ricardo. **Private credit bureaus and positive information sharing**: effects on credit cost?Brasília: Banco Central do Brasil, 2025. (Working Paper Series, n. 624). Disponível em: https://www.bcb.gov.br/content/publicacoes/WorkingPaperSeries/WP624v3.pdf. Acesso em: 21 fev. 2026.

PRINCE, Anya E. R.; SCHWARCZ, Daniel. **Proxy discrimination in the age of artificial intelligence and big data**. **Iowa Law Review**, Iowa City, v. 105, n. 3, p. 1257-1318, 2020. Disponível em: https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data. Acesso em: 30 abr. 2025.

CC BY NC SA

RUDIN, Cynthia. **Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence,** London, v. 1, p. 206-215, 2019. Disponível em: https://www.nature.com/articles/s42256-019-0048-x. Acesso em: 7 dez. 2025.

U.S. GOVERNMENT ACCOUNTABILITY OFFICE. **Artificial intelligence**: use and oversight in financial services. Washington, D.C.: GAO, 2025. (GAO-25-107197). Disponível em: https://www.gao.gov/assets/gao-25-107197.pdf. Acesso em: 20 abr. 2026.

WORLD BANK. **AI for risk-based supervision**. Washington, D.C.: World Bank, 2025. Disponível em: https://documents1.worldbank.org/curated/en/099021725190042788/pdf/P503970-2c607262-8673-4797-b5a6-52df7187e543.pdf. Acesso em: 20 abr. 2026.

WORLD BANK. **Brazil financial sector assessment**. Washington, D.C.: World Bank, 2024. Disponível em: https://documents1.worldbank.org/curated/en/099072425135514500/pdf/P181316-72eed651-b7c2-4137-a1f0-a00c3e1efede.pdf. Acesso em: 20 abr. 2026.