Algorithmic Credit and Financial Citizenship: institutional parameters for the use of artificial intelligence in credit underwriting

DOI: https://doi.org/10.58766/rpgbcb.v20.1.1345

Tayná dos Passos Carneiro*

Recebido/Received: 30/8/2026
Aprovado/Approved: 29/9/2026

Introduction. 1 From traditional credit to algorithmic credit: data, risk and financial inclusion. 1.1 Credit, informational asymmetry and risk assessment. 1.2 Digitalisation, alternative data and artificial intelligence. 1.3 Algorithmic legibility and the expansion of access to credit. 2 Algorithmic citizenship and the paradox of financial inclusion. 2.1 From financial citizenship to algorithmic citizenship. 2.2 From invisibility to hyperclassification. 2.3 Statistical accuracy and legal legitimacy. 3 Legal risks of algorithmic credit decision-making. 3.1 Algorithmic discrimination and proxies. 3.2 Opacity and explainability. 3.3 Contestability of automated decisions. 3.4 Human oversight and automation risk. 4 Institutional parameters for algorithmic credit underwriting. 4.1 Data governance. 4.2 Model governance and validation. 4.3 Impact assessment and non-discrimination. 4.4 Proportionate transparency. 4.5 Contestability and review. 4.6 Accountability and human oversight. 5 Regulation and supervision of algorithmic credit. 5.1 Regulatory fragmentation in Brazil. 5.2 International references and regulatory trends. Conclusion.

Abstract

Credit underwriting is progressively ceasing to be mediated by registry information and traditional banking history and coming to be mediated by algorithmic inference. The shift is ambivalent. The same informational infrastructure that makes it possible to assess profiles historically illegible to traditional models, and to that extent to broaden access, also expands the capacity to segment, price and exclude. The article argues that the central problem of algorithmic credit does not lie in the use of artificial intelligence as such, but in the absence of institutional structures capable of rendering its decisions auditable, intelligible, contestable and attributable. Methodologically, the study adopts a legal-dogmatic and legal-institutional approach, supported by interdisciplinary literature review, analysis of the applicable Brazilian rules and a functional examination of foreign and multilateral regulatory references. Drawing on the notion of financial citizenship adopted by the Banco Central do Brasil and on the literature on algorithmic citizenship, it proposes shifting the regulatory centre of gravity from explainability to effective contestability, and sets out a framework of six institutional parameters: data governance, model governance and validation, impact assessment and non-discrimination, proportionate transparency, contestability and review, and accountability with meaningful human oversight. It concludes that financial citizenship in algorithmically mediated markets is not measured by the quantitative expansion of access alone. It depends on the institutional quality of the decisions that define who obtains credit, on what terms and subject to what possibilities of understanding, review and challenge.

* *PhD in Law from the University of São Paulo (USP) and a Master's and Bachelor's degree in Law from the State University of Rio de Janeiro (UERJ), with a period of undergraduate study at Kobe University, Japan. She is a Professor at FGV Law São Paulo and a CodeX Affiliate at Stanford Law School. Her work focuses on financial regulation, banking law, artificial intelligence, payment systems, and digital assets. She serves as Vice President of the Artificial Intelligence Commission and Secretary-General of the Banking Law Commission of the Rio de Janeiro State Bar Association (OAB/RJ), and as an Advisor to the Brazilian Financial System Appeals Council (CRSFN). https://orcid.org/0009-0002-9532-558X*

Keywords: *algorithmic credit; financial citizenship; algorithmic discrimination; contestability; model governance; financial regulation.*

Crédito Algorítmico e Cidadania Financeira: parâmetros institucionais para o uso de inteligência artificial na concessão de crédito

Resumo

A concessão de crédito deixa progressivamente de ser mediada por informações cadastrais e histórico bancário tradicional para ser mediada por inferências algorítmicas. O deslocamento é ambivalente. A mesma infraestrutura informacional que permite avaliar perfis historicamente ilegíveis para os modelos tradicionais, e, nessa medida, ampliar o acesso, também amplia a capacidade de segmentar, precificar e excluir. O artigo sustenta que o problema central do crédito algorítmico não está no uso da inteligência artificial em si, e sim na ausência de estruturas institucionais capazes de tornar suas decisões auditáveis, inteligíveis, contestáveis e imputáveis. Quanto ao método, o trabalho adota abordagem jurídico-dogmática e jurídico-institucional, apoiada em revisão bibliográfica interdisciplinar, em análise da normativa brasileira aplicável e no exame funcional de referências regulatórias estrangeiras e multilaterais. A partir da noção de cidadania financeira adotada pelo Banco Central do Brasil e da literatura sobre cidadania algorítmica, propõe-se deslocar o eixo regulatório da explicabilidade para a contestabilidade efetiva, e apresenta-se um arranjo de seis parâmetros institucionais: governança de dados, governança e validação de modelos, avaliação de impacto e não discriminação, transparência proporcional, contestabilidade e revisão, e *accountability* com supervisão humana significativa. Conclui-se que a cidadania financeira em mercados mediados por algoritmos não se mede pela ampliação quantitativa do acesso. Ela depende da qualidade institucional das decisões que definem quem acessa crédito, em quais condições e mediante quais possibilidades de compreensão, revisão e contestação.

Palavras-chave: crédito algorítmico; cidadania financeira; discriminação algorítmica; contestabilidade; governança de modelos; regulação financeira.

Introduction

The informational infrastructure of Brazilian credit has changed structurally over the past decade. The digitalisation of financial services, the expansion of digital credit, the implementation of Open Finance, the growing availability of transactional data and the application of machine learning models to risk assessment have altered the speed at which credit is granted and, above all, the raw material of the credit decision. Motta and Rosa (2022, p. 137) observe that Big Data supplies the conditions for the raw volume of information produced within the financial system to be analysed through algorithmic processing, making it possible to identify behavioural patterns and to formulate risk predictions that are converted into customised financial products.

The shift can be summarised as follows: credit is progressively ceasing to be a relationship mediated by registry information and traditional banking history and becoming a relationship mediated by algorithmic inference. Digitalisation does not merely accelerate underwriting. It changes what can be known about the borrower. And if what can be known changes, so does the set of people the system is able to assess, together with the set of reasons for which it may refuse them.

From this the problem follows. Individuals previously excluded for informational insufficiency may be brought into the market by new risk models. The 2018 Financial Citizenship Report of the Banco Central do Brasil recorded that, among the population earning up to one monthly minimum wage, only 11% were credit borrowers, the lowest figure among all income brackets, and that these people are to be expected to face greater difficulty in obtaining credit because of the difficulty of proving income and offering collateral (Banco Central do Brasil, 2018, p. 26). The same report indicates that obtaining information by alternative means on the payment history of this population may be one of the solutions for broadening access (Banco Central do Brasil, 2018, p. 59). The figure dates from 2018 and should be read as a snapshot of that moment. Later editions of the report changed the methodological framing of the access indicator: the 2021 edition replaced the minimum-wage bracket with brackets denominated in reais, and the 2025 edition adopted the Cadastro Único as the criterion for delimiting the low-income population, so that none of them offers a directly comparable series (Banco Central do Brasil, 2021; Banco Central do Brasil, 2025). But the infrastructure capable of broadening financial inclusion is the same infrastructure capable of producing sophisticated mechanisms of discrimination, opacity and decisional asymmetry. The ambivalence is constitutive. More data means, at one and the same time, more capacity to include and more capacity to classify, segment and exclude.

The research question is the following: what institutional parameters are necessary for the use of artificial intelligence in credit underwriting to contribute to financial citizenship without reproducing or deepening forms of exclusion and decisional asymmetry?

The hypothesis advanced is that the central problem of algorithmic credit does not lie in the use of artificial intelligence as such, but in the absence of institutional structures capable of rendering its decisions auditable, intelligible, contestable and attributable. The legal legitimacy of the automated credit decision is not a function of its statistical accuracy. It is a function of the institutional architecture that surrounds it.

The study adopts a legal-dogmatic and legal-institutional approach. It draws on an interdisciplinary literature review spanning information economics, the sociology of economic classification, the empirical literature on alternative data and on machine learning applied to credit risk, and the legal literature on algorithmic discrimination and contestability. To this is added the analysis of the applicable Brazilian rules. International references are approached in a functional key, and not as an exhaustive comparative study: the aim is to extract reproducible regulatory functions, not to reconstruct the legislative history of each jurisdiction. The selection followed three cumulative criteria: direct bearing on credit decisions mediated by automated systems; the existence of a normative instrument or of supervisory guidance in force and publicly accessible; and representativeness of mutually distinct regulatory models, namely horizontal regulation by degree of risk, in the European Union, prudential discipline of the model life cycle, in the United Kingdom, non-binding supervisory guidance, in the United States, and multilateral systematisation, at the Bank for International Settlements. Jurisdictions that, at the date of the research, had no instrument in force applicable to credit underwriting were left out.

The text is organised in five movements. Section 1 reconstructs the path from traditional credit to algorithmic credit. Section 2 develops the category of algorithmic citizenship and formulates the paradox of digital financial inclusion. Section 3 examines four legal risks specific to the automated decision. Section 4, the propositional core of the study, sets out six institutional parameters. Section 5 discusses the Brazilian regulatory framework and the relevant international trends.

CC BY NC SA

1 From traditional credit to algorithmic credit: data, risk and financial inclusion

This section reconstructs the path that leads from credit mediated by registry information to credit mediated by inference. It starts from the informational problem at the origin of rationing and shows how digitalisation redefines who the system is able to assess.

1.1 Credit, informational asymmetry and risk assessment

Any analysis of algorithmic credit has to begin before artificial intelligence. Credit is a problem of decision under informational uncertainty. The lender must answer who the borrower is, what their capacity to pay is, what their probability of default is and what rate remunerates the risk assumed. And it must do so holding information systematically inferior to that held by the borrower.

Stiglitz and Weiss (1981) showed that asymmetry is not marginal noise. It is a structuring feature of how the market works. The authors argue that the interest rate charged by a bank may itself affect the riskiness of the loan portfolio, either by sorting potential borrowers, the adverse selection effect, or by changing the actions of borrowers already contracted, the incentive effect (Stiglitz; Weiss, 1981, p. 393). The analytical consequence is decisive: in equilibrium, the credit market may ration quantity instead of adjusting price. Credit restriction does not follow only from the cost of money. It also follows from the difficulty of measuring individual risk.

This diagnosis explains why the classical mitigation instruments perform an economic function, and not a merely bureaucratic one. Credit history, positive and negative registries, in rem and personal guarantees and the requirement of formal proof of income are technologies for reducing informational asymmetry. It also explains why they produce systematic exclusion. Someone with no history, no guarantee and no formally verifiable income is not assessed as a bad risk. They are, more precisely, not assessed. To this condition corresponds the Brazilian figure already cited, a credit penetration of 11% in the income bracket up to one minimum wage, measured in 2018 (Banco Central do Brasil, 2018, p. 26).

The very definition of the object of prediction is an institutional construction. Barocas and Selbst (2016, p. 679) observe that the predicted probability of failing to make a given number of payments is not a self-evident answer to the question of how to extend credit to consumers. Unlike fraud or spam, creditworthiness is an artefact of the definition of the problem itself, because there is no way of measuring it directly: the notion depends on the particular way in which the credit industry has built the system of underwriting and payment. The observation seems technical. It has a direct legal consequence, and will be taken up again below. If the target of the prediction is a choice, then it is a choice open to scrutiny.

1.2 Digitalisation, alternative data and artificial intelligence

The digitalisation of financial services alters the informational infrastructure described above. Under the previous paradigm, information remained concentrated in the few institutions holding the prior relationship with the customer, in a regime of closed data. Motta and Rosa (2022, p. 138) observe that this concentration constituted a barrier to the entry of competitors willing to draw on new data processing technologies. Open Finance breaks the arrangement by instituting the sharing of data and services among participating institutions with the customer's consent.

This is not merely a matter of more data on the same variables. It is a matter of variables of another nature: granular transactional flows, payment behaviour in continuing services, platform data and, at the limit, digital traces produced incidentally. The empirical evidence on the predictive power of such data is robust. Berg et al. (2020) analysed some 250,000 observations from a German e-commerce company and showed that a model built exclusively from simple variables of the so-called digital footprint reaches an area under the curve (AUC) of 69.6%, higher than the AUC of 68.3% obtained with the credit bureau score alone (Berg et al., 2020, p. 2847). The variables are trivial from the standpoint of collection: device type, operating system, e-mail provider, access channel, time of purchase and characteristics of the e-mail address supplied.

The mechanism matters for the argument of this article. The authors record that orders originating from the Android operating system showed a default rate of 1.79%, almost double that observed on iOS systems, 1.07% (Berg et al., 2020, pp. 2856-2857). And they explain that such variables provide accessible proxies for a person's economic status, in the absence of private information and of income data that is difficult to collect (Berg et al., 2020, p. 2854). The predictive power of the digital footprint derives, in relevant part, from its capacity to approximate socio-economic variables that were not declared. In many legal orders, those variables could not be used directly either.

1.3 Algorithmic legibility and the expansion of access to credit

The inclusive consequence of the shift is real and should not be minimised. Berg et al. (2020, p. 2882) show that the discriminatory power of the digital footprint for customers without a bureau score, the unscorable customers, whose history is insufficient for the calculation of a traditional score, matches the discriminatory power observed for customers with a score: an AUC of 72.2% against 69.6% within the sample and of 68.4% against 68.8% outside it (Berg et al., 2020, pp. 2882 and 2884). The authors conclude that the digital footprint has the potential to widen the financial inclusion of part of the roughly two billion adults who currently lack access to credit (Berg et al., 2020, p. 2849).

The phenomenon deserves a name of its own. Let us call financial legibility the condition of an individual being assessable by the informational instruments available to a given credit system. Under the traditional regime, profiles marked by informality, irregular income, scant documentation and the absence of guarantees were, in a precise sense, illegible: they produced no signals that the models knew how to read. Digitalisation converts traces that were previously irrelevant into processable risk signals. The individual invisible to the traditional model becomes visible and, therefore, assessable, priceable and possibly included.

The formulation recovers, in a contemporary register, a settled sociological intuition. Fourcade and Healy (2013, p. 559) hold that, in the neoliberal era, market institutions increasingly use actuarial techniques to slice and sort individuals into classification situations that shape their life chances, and that the United States credit market operated simultaneously as a levelling force, by incorporating previously excluded groups, and as a condenser of new forms of social difference. The authors conclude that classification situations may have become the engine of modern class situations (Fourcade; Healy, 2013, p. 559). Widening access and producing new differences are not opposing movements. They are the same movement seen from two angles.

2 Algorithmic citizenship and the paradox of financial inclusion

With the economic and technological basis established, the normative category that allows it to be assessed must now be built. This section starts from the financial citizenship adopted by the Banco Central do Brasil, transposes the literature on algorithmic citizenship to the domain of credit and formulates the paradox that organises the argument.

CC BY NC SA

2.1 From financial citizenship to algorithmic citizenship

The normative starting point is the category of financial citizenship built by the Banco Central do Brasil. According to the Financial Citizenship Report, the development of financial citizenship takes place through a context of financial inclusion, financial education, protection of the consumer of financial services and participation in the dialogue about the financial system (Banco Central do Brasil, 2018, p. 8). To be a financial citizen in full, the document continues, includes having access to services suited to the individual's needs, developing capabilities to manage resources, having a business environment that generates trust, with simple information and dispute resolution mechanisms, and having channels through which to take part in the debate on how the financial system works (Banco Central do Brasil, 2018, p. 8).

The definition is institutionally dense, and not rhetorical. In detailing the indicators of the inclusion dimension, the report sets out four characteristics: suitability, understood as the adequacy of the options to the capabilities and interests of the various profiles; proportionality of costs, documentary requirements and collateral requirements in the face of the risks and complexity of the service; convenient delivery; and equity, defined as dignified and equitable treatment for every citizen, without prejudice as to origin, race, gender, age or any other form of discrimination (Banco Central do Brasil, 2018, p. 51). In the same passage, the report adopts the World Bank formulation according to which the protection of the consumer of financial services seeks to ensure that consumers receive information enabling them to take informed decisions, are not subject to unfair or deceptive practices and have access to mechanisms and to the lodging of appeals in order to resolve disputes (Banco Central do Brasil, 2018, p. 51).

The consequence is underused in the Brazilian debate. Non-discrimination and effective mechanisms of recourse are not external additions to the regulatory concept of financial citizenship. They are elements internal to it, as formulated by the monetary authority itself. A system that increases the number of credit agreements while degrading equity of treatment or the possibility of recourse does not widen financial citizenship. It displaces it.

The expression algorithmic citizenship is not a creation of this work and has a theoretical genealogy of its own, which needs to be made explicit so that it is not used as a label. Cheney-Lippold (2016) coined the notion of jus algoritmi in analysing how the United States National Security Agency came to determine the status of citizen or foreigner from the processing of communications data, giving the term the sense of citizenship rights according to algorithmic logic (Cheney-Lippold, 2016, p. 1722). It is, in the author's words, a citizenship of technical requirement, one that can neither be intentionally practised nor remain functionally stable (Cheney-Lippold, 2016, p. 1723). Tomasello (2023, p. 2) records that notions such as datafied citizenship and algorithmic citizenship were coined to describe a reality in which identities are assigned by continuous data processing.

The possible originality does not lie in naming the concept. It lies in transposing it critically to the specific domain of financial citizenship. The following operational definition is proposed: algorithmic citizenship, in the context of credit, refers to the institutional conditions necessary for individuals to remain subjects of rights in environments in which relevant economic decisions are increasingly mediated by automated systems. With growing algorithmic intermediation, part of the conditions of financial citizenship set out by the Central Bank comes to depend on how individuals are represented by data, how they are classified, what inferences are drawn about them, how decisions are explained to them and what mechanisms exist for contesting them.

Cheney-Lippold's critique of jus algoritmi is instructive on the last point. The author observes that being unaware of one's own status as a member of the political community frustrates the reflexivity required by the traditional type of citizenship: when the user is secretly seen as a foreigner, no answer or judgement endowed with power returns to them as feedback (Cheney-Lippold, 2016, p. 1736). And he adds that the user's actions are weighed by a computer and by an analyst while the quantitative definitions of their algorithmic citizenship remain hidden, resources, rights and life possibilities being distributed on the basis of statistical interpretations of those actions (Cheney-Lippold, 2016, p. 1738). Replace intelligence agency with financial institution and the status of foreigner with risk class. What results is the exact structure of the problem of algorithmic credit.

2.2 From invisibility to hyperclassification

It is now possible to state the paradox that organises this article. Digitalisation may displace the problem of financial exclusion: from a system that excluded because it knew little about certain individuals to a system capable of excluding them because it believes it knows a great deal about them.

The shift is qualitative. Exclusion by informational insufficiency is undifferentiated: the system cannot tell individuals apart and, as a precaution, refuses or raises the price. Exclusion by hyperclassification is targeted. The system segments finely, assigns probabilistic profiles, personalises prices and distinguishes statistically correlated groups with increasing granularity. The first regime produces under-inclusion. The second may produce adverse inclusion, with access granted on systematically worse terms for certain groups.

The empirical evidence sustains the concern. Fuster et al. (2022, p. 5) examined the distributive effects of replacing traditional models with machine learning in the United States mortgage market. They concluded that Black and Hispanic borrowers are disproportionately less likely to benefit from the introduction of machine learning, and that, in a simple equilibrium model of the credit market, machine learning increases rate disparity between groups and within them. The authors record that the distributive effects of a superior statistical technology may stem from greater flexibility in uncovering structural relationships or from the triangulation of excluded characteristics (Fuster et al., 2022, p. 5). Triangulation is precisely the proxy mechanism discussed below.

None of this means that algorithmic intermediation is necessarily worse than human intermediation. Bartlett et al. (2022, p. 30) document, in the same market, that lenders charge Latino and African-American borrowers 7.9 and 3.6 basis points more on purchase and refinance mortgages, and that fintech algorithms also discriminate, though 40% less than lenders operating face to face. The finding is doubly relevant. It confirms the persistence of discrimination in the digital environment. And it disallows the simplistic narrative that automation is, in itself, the problem. The problem lies in the institutional architecture, not in the technological substrate.

There is a further risk, of another nature. Berg et al. (2018, p. 26) warn that, if the digital footprint comes to determine access to credit, consumers will tend to adapt their behaviour, and that a world of conformity, in which consumers fear expressing their individual personality and act with the permanent desire to project a positive image before third parties, is not the model of society most people have in mind. The externality escapes the traditional categories of consumer law and touches the free development of personality, a legal interest expressly protected by Article 1 of Law No. 13,709 of 14 August 2018, the Brazilian General Data Protection Law (LGPD).

2.3 Statistical accuracy and legal legitimacy

The conceptual core of the argument condenses into a single proposition: statistical accuracy is not equivalent to legal legitimacy. A model may predict correctly and still produce legally problematic differentiation, indirect discrimination, disproportionate treatment, exclusion grounded in proxies and the practical impossibility of contestation.

Barocas and Selbst (2016, p. 677) put the point precisely. By definition, data mining is always a form of statistical discrimination, and therefore apparently rational, since its purpose is to provide a rational basis for distinguishing between individuals and for attributing to the individual case the qualities of those who are statistically similar to it. The observation shifts the axis of the debate. The problem is not that the model discriminates, in the technical sense of distinguishing. The problem is to determine which distinctions the legal order admits, on what grounds and under what safeguards.

The conclusion the authors reach is still more uncomfortable for the thesis that better prediction solves the distributive problem. If an attribute is unequally distributed in the population, a more accurate form of data mining will be more likely to capture that distribution; better data and more attributes will only come closer to the exact

CC BY NC SA

extent of existing inequality (Barocas; Selbst, 2016, p. 692). In a structurally unequal society, predictive improvement tends to make inequality more visible to the model and, absent institutional constraints, more operative in the decision.

The formulation rules out two current simplifications. The first is the mantra according to which biased data generates biased algorithms, true but intellectually thin, because it suggests that correcting the data would solve the problem. The second is the inverse assumption, that technically correct data produces neutral decisions. Barocas and Selbst (2016, p. 691) show that discrimination may occur even when protected criteria have been removed from the dataset, the data is free of latent prejudice, the attributes are especially granular and diverse, and the sole objective is to maximise classification or prediction accuracy. The legally decisive point is this: technically correct data may reproduce unequal social structures without any agent having acted with discriminatory intent.

3 Legal risks of algorithmic credit decision-making

The paradox unfolds into four legal risks: discrimination through proxies, opacity, the absence of effective contestability, and merely formal human oversight. Each calls for an institutional response of its own, and from this follows the arrangement proposed in section 4.

3.1 Algorithmic discrimination and proxies

The first risk is discrimination. Three legally distinct situations must be separated: the explicit use of a protected attribute as a model variable; indirect discrimination, in which an apparently neutral criterion produces a disproportionate adverse effect on a protected group; and the use of proxies, in which correlated variables reconstruct the protected attribute.

Brazilian law already prohibits the first of these in the credit domain. Law No. 12,414 of 9 June 2011, the Positive Credit Registry Law, prohibits, in Article 3(3)(I) and (II), the recording of excessive information, understood as information not connected to the assessment of consumer credit risk, and of sensitive information, understood as information pertaining to social and ethnic origin, health, genetic information, sexual orientation and political, religious and philosophical convictions. Law No. 13,709/2018 reinforces the prohibition by enshrining, in Article 6(IX), the principle of non-discrimination, barring the processing of personal data for unlawful or abusive discriminatory purposes.

The difficulty lies in the two remaining situations. Barocas and Selbst (2016, p. 691) identify the mechanism they call redundant encodings: cases in which membership in a protected class is itself encoded in other data, because a given item of information, or given values of that information, is highly correlated with membership in the group. In such cases, the criteria genuinely relevant to a rational and well-informed decision also function as reliable indicators of class membership. Fuster et al. (2022, p. 5) name the same phenomenon triangulation of excluded characteristics.

The empirical literature on alternative data makes the risk concrete. If digital footprint variables function as accessible proxies for a person's economic status (Berg et al., 2020, p. 2854), and if the operating system of the device predicts income quartile, then the boundary between a legitimate risk variable and the indirect reconstruction of a socially structured attribute becomes thin. The authors themselves anticipate the regulatory problem when they observe that regulators tend to scrutinise closely whether the digital footprint functions as a proxy for information whose use is prohibited by fair lending laws (Berg et al., 2020, pp. 2849 and 2886).

In Brazil, Amparo and Prado (2024) present a mapping of the rules prohibiting discrimination in access to credit and assess the discriminatory risks associated with the use of data in credit underwriting, proposing a classification into three categories of racial discrimination risk: low risk, for data whose use is in principle acceptable; high risk, for

data in principle unacceptable for credit assessment; and uncertain risk, for data situated in a grey area. The proposal is methodologically relevant. It shifts the analysis from the isolated data point to the examination of its function within the decision-making structure. And it recognises a grey area that demands an institutional decision, not merely a technical one.

Motta and Rosa (2022, p. 148) reach a convergent conclusion. They warn that, depending on the information feeding the system, algorithmic insights may lead to discriminatory outcomes in relation to race, gender or any other condition of the data subject, without the operators having programmed the system to that end, and that such an outcome offends the constitutional order even where the machine's conclusion was reached on the basis of accurate data. This is precisely the case in which accuracy does not redeem the decision.

3.2 Opacity and explainability

The second risk is opacity. It presents itself at three levels, and each calls for a distinct institutional response. There is technical opacity, arising from the intrinsic complexity of the model. There is organisational opacity, arising from insufficient documentation, the fragmentation of internal responsibilities or the engagement of a third-party vendor. And there is informational opacity, arising from the insufficiency of the explanation offered to the consumer, even where the institution holds the information internally.

Crisanto et al. (2024, p. 23) summarise the problem at the supervisory level. As artificial intelligence systems become more complex, they frequently achieve superior performance at the expense of explainability, so that striking the right balance between performance and explainability is one of the main challenges facing financial institutions, particularly in credit and insurance underwriting. In the same passage, the authors record that the lack of transparency as to the way such systems produce credit decisions raises concerns about compliance with consumer protection and model risk management requirements, since consumer regulation typically requires the institution to inform the customer of the principal reasons for a denial.

In Brazilian law, the matter is anchored in Article 20 of Law No. 13,709/2018, which guarantees the data subject the right to request the review of decisions taken solely on the basis of automated processing of personal data that affect their interests, including decisions intended to define their credit profile. Article 20(1) requires the controller to provide, whenever requested, clear and adequate information on the criteria and procedures used in the automated decision, subject to commercial and industrial secrets.

Motta and Rosa (2022, p. 149) identify the weakness of this formulation with precision. Their criticism does not fall on the legal appropriateness of safeguarding commercial and industrial secrets, a protection established by the Brazilian statute and not expressly provided for in the European General Data Protection Regulation. It falls on the open and unrestricted use of the expression, which may become a problem in practice. The authors observe that, given the breadth of the expression, the processing agent may respond to a request for explanation by asserting that the automated scoring process is covered by trade secrecy, and that it often in fact is (Motta; Rosa, 2022, p. 149). To this is added the black-box effect, which may render it practically unfeasible to demonstrate compliance with both Open Finance and the General Data Protection Law (Motta; Rosa, 2022, p. 150).

The point to be sustained is that explainability, necessary though it is, has occupied excessive space in the Brazilian debate. Explainability techniques, model cards and contributing-factor reports (reason codes) are useful instruments. None of them answers the legally decisive question: what the consumer can do with the explanation received.

CC BY NC SA

3.3 Contestability of automated decisions

We arrive at the central shift of this article. The legally relevant criterion should not be explainability. It should be effective contestability.

Recent literature offers a robust foundation for the shift. Alfrink et al. (2023, p. 613) define contestable by design systems as systems that are open and responsive to human intervention throughout their life cycle. Responsiveness is not confined to the moment following the automated decision. It reaches the design and development of the system as well. Systematising the literature, drawing on Almada (2019) and on Sarra (2020 apud Alfrink et al., 2023, p. 615), the authors record that to contest is to engage with the substance of the decision itself, which requires more than voicing an opinion: it requires an articulated act of defence, which in turn requires arguments, and arguments require information (Alfrink et al., 2023, p. 615). They conclude that contestability requires explainability, with an explanation that includes both the description of the how and the justification of the why, and that it is not reducible to it, because it requires a procedural relationship: a human in the loop is insufficient if there is no possibility of dialectical exchange between the subject of the decision and the human controller (Alfrink et al., 2023, p. 615).

The operational content of contestability breaks down into five faculties. The consumer must be able to understand the nature of the decision, to identify the elements that determined it, to submit corrective information, to challenge errors of fact and of classification, and to obtain meaningful review. Alfrink et al. (2023, p. 624) argue that, for genuine contestation to exist, it is not enough that the subject be able to express their point of view, a dialectical exchange between subject and controller being necessary, and that contestability must include the right to a new decision, to compensation for the harm caused or to the reversal of the decision.

The requirement is not foreign to Brazilian law. It is found there in embryonic form. Súmula No. 550 of the Superior Court of Justice, a binding-style summary of settled case law, establishes that the use of credit scoring, a statistical method of risk assessment that does not constitute a database, dispenses with the consumer's consent, the consumer retaining the right to request clarification regarding the personal information assessed and the sources of the data considered in the respective calculation. In its judgment of Special Appeal (Recurso Especial) No. 1,419,697/RS, submitted to the repetitive appeals regime and the origin of Theme No. 710, the docket entry under which the resulting binding precedent is catalogued, the Second Panel held that, although the consent of the consumer consulted is unnecessary, clarification must be provided to them, if requested, as to the sources of the data considered and the personal information assessed, and that disregard for the statutory limits on the use of the system, notably through the use of excessive or sensitive information prohibited by Article 3(3)(I) and (II) of Law No. 12,414/2011, as well as the improper denial of credit through the use of incorrect or outdated data, may give rise to compensation.

Law No. 12,414/2011 guarantees the registered individual, in Article 5(IV), the right to know the principal elements and criteria considered in the risk assessment, subject to trade secrecy, and, in item VI of the same article, the right to request from the enquiring party the review of a decision taken exclusively by automated means. The Consumer Protection Code, in Article 43(1), requires consumer records and data to be objective, clear, truthful and in easily comprehensible language, and, in Article 43(3), guarantees the immediate correction of inaccurate data.

The normative material exists. What is missing is its articulation around a unifying criterion, and that is the contribution intended here. Explanation is a necessary condition, not a sufficient one. What gives it its point is the possibility of contesting. A right to explanation without a correlative right of correction and of substantive review turns into ritualistic transparency: the duty to inform is discharged and nothing changes in the legal position of the person affected.

3.4 Human oversight and automation risk

The fourth risk is that of merely formal human intervention. Crisanto et al. (2024, p. 18) record that regulatory guidance on artificial intelligence emphasises the role of human intervention in minimising the risk that automated decisions produce harmful outcomes, above all where those outcomes have a potentially significant impact on customers, and that three concepts stand out: human-in-the-loop, corresponding to human intervention in the decision cycle; human-on-the-loop, corresponding to intervention during the design cycle and in subsequent reviews; and human-in-control, corresponding to human primacy in critical decision-making.

The taxonomy should not be treated as decorative. The legally relevant point is that human intervention adds legitimacy only where there is effective capacity for review, adequate information and decisional autonomy. Alfrink et al. (2023, p. 623), synthesising Almada (2019) and Brkan (2019 apud Alfrink et al., 2023, p. 623), put the warning directly: the mere presence of a human at the end of the decision chain, the so-called human in the loop, may not constitute a sufficient safeguard against machine error if the human controllers lack the authority or the capacity to base their final decision on more information than that supplied by the system. By extension, the human controllers who respond to requests for intervention must have the authority and the capacity effectively to alter prior decisions.

Without these conditions, human oversight degenerates into a human signature on an algorithmic decision, a phenomenon the English-language literature designates as rubber stamping. The legal consequence is serious. The formality of the intervention may be invoked to exclude the application of Article 20 of Law No. 13,709/2018, on the argument that the decision was not taken solely on the basis of automated processing, hollowing out the guarantee the provision was meant to establish. The same authors record that human intervention may itself be biased and lead to worse outcomes than those of the fully automated procedure, which is why comparative performance measures between the human-controlled procedure and the automated one should be adopted (Alfrink et al., 2023, p. 623). Meaningful human oversight is measured. Providing for it in a rule is not enough.

4 Institutional parameters for algorithmic credit underwriting

With the diagnosis established, the proposal follows. The legitimacy of algorithmic credit depends on a governance architecture articulated around six institutional parameters, whose adoption must observe proportionality in relation to the size of the institution, model materiality and the potential impact on the customer.

4.1 Data governance

The first parameter bears on the input. It requires quality, relevance, currency, traceability of origin and fitness for purpose of the data employed. It also requires a specific assessment of their potential to operate as proxies for protected attributes, and a dedicated regime for alternative data.

The normative anchoring in Brazil is consolidated. Joint Resolution No. 1 of 4 May 2020, issued by the National Monetary Council and the Banco Central do Brasil, in regulating Open Finance, lists among its objectives the promotion of financial citizenship, in Article 3(IV), and sets out, in Article 4, among the principles to be observed by participating institutions, transparency, in item I, data quality, in item III, and non-discriminatory treatment, in item IV, requiring them to act ethically and responsibly in the pursuit of those objectives and principles (Conselho Monetário Nacional; Banco Central do Brasil, 2020). To the same effect, Motta and Rosa (2022, p. 140). Law No. 12,414/2011 prohibits excessive and sensitive information. Law No. 13,709/2018 imposes the principles of adequacy, necessity, data quality and non-discrimination.

CC BY NC SA

Crisanto et al. (2024, p. 24) record that the expectation of transparency, in its consumer dimension, and its connection with expectations of fairness heighten the role of data governance. Documenting the data used is important both for explaining the resulting decisions to the customer and for assessing which supplementary data provided by the customer are relevant. Institutions must assess whether input data are biased and adopt policies ensuring lawful, ethical and secure collection, storage, processing and sharing.

4.2 Model governance and validation

The second parameter bears on the model. It comprises documentation of design and assumptions, independent validation, prior testing, continuous performance monitoring, version control, lifecycle management, periodic review at a cadence proportionate to materiality and a clear definition of model ownership.

The central notion is simple. Credit models are decision-making infrastructures, not mere technological tools. That characterisation has consequences: they are subject to the governance regime proper to structures that produce legal effects on third parties, not to the regime governing software procurement.

The most consolidated international reference is the discipline of model risk management. In the United Kingdom, Supervisory Statement SS1/23 of the Prudential Regulation Authority, published in May 2023, effective from 17 May 2024 and consolidated in a new version on 23 April 2026, organises supervisory expectations into five principles: model risk identification and classification; governance; development, implementation and use; independent validation; and risk mitigants. Two points are of direct interest to algorithmic credit. Principle 1.3(c) provides that the assessment of model complexity may also consider, where necessary and above all in the use of newly advanced approaches or technologies, factors connected to the use of alternative and unstructured data and to measures of interpretability, explainability and transparency, as well as the potential for design or data bias. Principle 3.2(b) provides that the development process should ensure the absence of inappropriate bias in the data and the compliance of their use with data privacy and other relevant data regulations. The PRA already treats alternative data and the potential for bias as model risk classification factors, and not as a matter external to prudential discipline.

In the United States, on 17 April 2026, the Federal Reserve, the Office of the Comptroller of the Currency and the Federal Deposit Insurance Corporation jointly issued revised guidance on model risk management, conveyed by Supervisory Letter SR 26-2, which expressly replaces Supervisory Letters SR 11-7, of 2011, and SR 21-8, of 2021, and adopts a risk-based approach, calibrated to the model risk profile and to the size and complexity of the banking organisation's operations.

Two observations counsel caution in transplanting this reference. The first is that SR 26-2 is expressly non-binding and states that non-compliance will not, in itself, result in supervisory criticism, being addressed above all to banking organisations with more than USD 30 billion in total assets. The second matters more to the object of this article: the guidance excludes generative and agentic artificial intelligence from its scope, referring their governance to the institution's general risk management principles. There is a mismatch here worth recording. At the very moment when financial authorities identify model risk management frameworks as the primary instrument for mitigating artificial intelligence risks, including opacity and lack of explainability (Crisanto et al., 2024, p. 23), the leading United States guidance on the subject leaves the most recent technological frontier outside its scope. This is not a criticism of the regulatory choice, which is justified by prudence in the face of rapidly evolving technology. It is an observation that the Brazilian architecture cannot presume that importing a model risk framework will resolve, on its own, the problem of algorithmic credit.

It remains to ask whether this anchoring needs to be imported. CMN Resolution No. 4,557 of 23 February 2017 already contains the central elements of model governance, even though it does not use the expression model risk. Article 9 requires periodic assessment of the models used in risk management as to the adequacy and robustness of their

assumptions and methodologies and as to their performance, including the comparison between estimated and observed losses. The sole paragraph of that article prohibits such assessment from being conducted by the unit responsible for developing the model or by a risk-taking unit, enshrining the independence of the assessment function. Article 12(V)(f) requires, within the stress-testing programme, documentation of the methodological limitations connected to model selection, to their assumptions and to the databases used. And Article 51(IV) assigns to the board of directors, the risk committee, the chief risk officer and the executive board the duty to understand the limitations and uncertainties relating to the models, even when developed by third parties. The domestic normative basis therefore exists. What is missing is its explicit extension to the models that determine the customer's access to credit, which do not, as a rule, fall within the perimeter of the prudential capital models to which the rule was primarily addressed.

4.3 Impact assessment and non-discrimination

The third parameter requires prior and continuous testing of performance disaggregated by relevant groups, false positive and false negative rates, identification of proxies, observed adverse effects and model drift over time.

The distinction proposed by Crisanto et al. (2024, p. 18) is operationally useful here. The authors identify two dimensions of fairness present in regulatory guidance. Distributive fairness concerns the fairness of the outcomes produced by decisions, that is, non-discrimination, and is the most frequently cited and at the same time the most difficult to measure and to achieve. Procedural fairness concerns the fairness of the decision-making process, operationalised through the concepts of external transparency and external accountability. The institutional architecture must address both. Addressing only the second turns impact assessment into a documentary exercise.

Comparative law offers a mature institutional model. Regulation (EU) 2024/1689 classifies as high risk, in Annex III, point 5(b), artificial intelligence systems intended to evaluate the creditworthiness of natural persons or to establish their credit score, with the exception of systems used for the detection of financial fraud. Article 27(1) of the same Regulation extends the obligation to carry out a fundamental rights impact assessment to bodies governed by public law, to private entities providing public services and to the deployers of the high-risk systems referred to in Annex III, point 5(b) and (c). Private banks and fintechs that employ creditworthiness assessment systems will become subject, in the European Union, to a fundamental rights impact assessment prior to putting the system into service. The timetable, however, has changed. Regulation (EU) 2026/1744, adopted on 8 July 2026 and in force since 27 July 2026, amended Article 113 of Regulation (EU) 2024/1689 and postponed the application of Sections 1 to 3 of Chapter III, in which Article 27 sits, from 2 August 2026 to 2 December 2027 for the high-risk systems in Annex III, referred to in Article 6(2), and from 2 August 2027 to 2 August 2028 for those in Annex I, referred to in Article 6(1). The obligation exists, and the regulatory function to be drawn from it remains the same. Its enforceability for creditworthiness assessment, however, begins only in December 2027.[1]

This is the most relevant regulatory function to be drawn from the European experience. Not the risk taxonomy itself, but the requirement that the impact assessment be documented, prior and capable of being asserted before the authority, and that it fall upon whoever operates the system, and not only upon whoever develops it.

1 *Regulation (EU) 2026/1744 postponed, in its literal terms, only Sections 1 to 3 of Chapter III, in which Article 27 sits. Article 86, which secures the right to obtain explanations of the individual decision, belongs to Chapter IX and was not mentioned in the postponing provision, remaining enforceable since 2 August 2026. Since Article 86 refers, however, to decisions taken on the basis of high-risk systems referred to in Annex III, and since those systems become subject to the high-risk regime only on 2 December 2027, an interval subsists in which a right to explanation exists without the duties of documentation, record-keeping, human oversight and impact assessment that would give it content being in force. The mismatch illustrates the thesis argued here: explanation, unaccompanied by the institutional structure that makes contestation possible, operates as a formal guarantee.*

CC BY NC SA

4.4 Proportionate transparency

The fourth parameter sets aside an unproductive aspiration: absolute transparency. Trade secret is a legally protected interest, safeguarded by Article 20(1) of Law No. 13,709/2018 and by Article 5(IV) of Law No. 12,414/2011. Crisanto et al. (2024, p. 26) observe that third-party models may not allow full visibility of proprietary information, and that requiring its disclosure would expose intellectual property and confidential business information, discouraging innovation.

The appropriate architecture does not choose between opacity and full disclosure. It distinguishes four planes with distinct extents. To the regulator, full transparency: data, code, assumptions, validation documentation and disaggregated metrics, subject to the authority's duty of confidentiality. Internally, transparency sufficient for senior management and the risk, compliance and audit functions to exercise effective challenge. To the consumer, functional transparency: the nature of the decision, the main elements considered and the path to contestation. To the public, aggregate transparency, with metrics and policies that allow social scrutiny without exposing trade secrets.

Alfrink et al. (2023, p. 628) offer the corresponding institutional design. Where the operators of the system have a legitimate claim to secrecy, third parties may act as trusted intermediaries to whom the sensitive information is disclosed, both for ex ante inspection of the system as a whole and for the ex post contestation of individual decisions. The solution is compatible with the Brazilian institutional design. Article 20(2) of Law No. 13,709/2018 already authorises the national authority to carry out an audit to verify discriminatory aspects in automated processing where the controller fails to provide information on the ground of observing commercial and industrial secrets. That provision, little explored today, is the instrument that makes it possible to reconcile trade secret and public scrutiny.

4.5 Contestability and review

The fifth parameter converts the discussion in section 3.3 into institutional design. It proposes the recognition of an operational, and not merely abstract, right to contest the decision, correct incorrect or outdated data, submit additional relevant elements, trigger substantive review and receive an adequate justification for the outcome of that review.

Two requirements are commonly omitted. The first is that the institution actually consider the supplementary information submitted by the customer. Crisanto et al. (2024, p. 20) record that financial institutions using artificial intelligence with potentially significant impact on customers should offer them channels to make enquiries, lodge appeals and request review of the decisions affecting them, and that both the European Regulation and the guidance of the Monetary Authority of Singapore provide that deployers have mechanisms to take into account verified and relevant supplementary data supplied by customers in the review of decisions. The second is that the review be accompanied by a justification. Communicating the outcome is not enough.

The parameter engages with Article 86(1) of Regulation (EU) 2024/1689, which grants a person affected by a decision taken on the basis of the output of a high-risk system listed in Annex III the right to obtain from the deployer clear and meaningful explanations of the role of the system in the decision-making procedure and of the main elements of the decision taken. The European formulation is notably close to Súmula No. 550 of the Superior Court of Justice, which grants the consumer the right to request clarification about the personal information assessed and the sources of the data considered in the calculation of the score. The convergence suggests that Brazil does not need to import the content of the right. It needs to build the procedure that makes it exercisable.

It is here that algorithmic citizenship acquires concreteness. If the distinguishing feature of that condition is the absence of feedback endowed with power for the classified subject (Cheney-Lippold, 2016, p. 1736), the circuit of effective contestation is the device that restores to that subject the position of a bearer of rights.

4.6 Accountability and human oversight

The sixth parameter answers the simplest and most avoided question: who answers for the model? The business area, risk management, compliance, the external supplier or the regulated institution?

The institutional answer must be unequivocal. Technological outsourcing does not entail the outsourcing of responsibility. Crisanto et al. (2024, p. 26) record that the regulatory principle whereby the board and senior management of financial institutions are ultimately responsible for any activities, functions, products or services provided by third parties applies equally to artificial intelligence. From this follows the duty to select third-party models appropriately and to validate them against the same standards applicable to internal models, and to agree contractual clauses requiring the supplier to provide evidence of fitness for the intended use, test results and information on the model's limitations and assumptions. Supervisory Statement SS1/23 adopts an equivalent solution in Principle 2.6(a): boards and senior management remain responsible for model risk even in outsourcing arrangements.

The same authors record that procedural fairness remains challenging in that arrangement. It is not clear that institutions can explain to customers decisions determined above all by third-party foundation models, rather than by the customisation they have carried out. And, if something goes wrong, they will be held liable regardless of whether they built or purchased the model (Crisanto et al., 2024, p. 26). The conclusion is direct. If responsibility lies with the regulated institution in any event, the framework must secure to it the contractual and informational means to exercise it.

As to human oversight, the parameter requires more than formal provision: named attribution of responsibility for the model, training of reviewers, access to information going beyond the output of the system, decision-making autonomy to overturn it and comparative measurement between the performance of the reviewed procedure and that of the fully automated one (Alfrink et al., 2023, p. 623). Crisanto et al. (2024, p. 2) summarise the corresponding supervisory expectation: the governance framework may need to specify the role of human intervention in order to minimise harmful outcomes of artificial intelligence systems.

5 Regulation and supervision of algorithmic credit

The parameters proposed presuppose a regulatory arrangement capable of sustaining them. This section examines the Brazilian framework, marked by overlapping regimes, and draws from the European, British, United States and multilateral experiences regulatory functions that can be transposed.

5.1 Regulatory fragmentation in Brazil

The argument of this section is that the Brazilian problem is not an absence of rules. It is fragmentation across distinct legal regimes, with different authorities, rationales and enforcement instruments bearing simultaneously on the same object.

Algorithmic credit is subject to at least six regulatory layers. Law No. 13,709/2018 governs the processing of personal data, the principles of non-discrimination and transparency and the right to review of automated decisions. Law No. 12,414/2011 regulates payment-history databases, prohibits excessive and sensitive information and secures the right to know the principal elements considered in the risk assessment. Law No. 8,078 of 11 September 1990 subjects consumer records and databases to duties of objectivity, clarity, truthfulness and correction. Joint Resolution No. 1/2020 of the CMN and the BCB establishes, within Open Finance, principles of data quality and of non-discriminatory treatment. CMN Resolution No. 4,949 of 30 September 2021 provides in Article 2 that institutions shall conduct their

CC BY NC SA

activities in dealings with customers and users in observance of the principles of ethics, responsibility, transparency and diligence, and requires in Article 3(II) that they afford fair and equitable treatment to customers and users, taking into account their relationship profiles and associated vulnerabilities. BCB Resolution No. 155 of 14 October 2021 established, for consortium administrators and payment institutions, principles and procedures similar to those of CMN Resolution No. 4,949/2021, with the declared purpose of harmonising that regime with the one applicable to financial institutions, subject to the specific features of each segment. Its scope was subsequently widened to reach securities brokers and dealers, foreign exchange brokers and virtual asset service providers. To this set is added the legislative debate now under way on the regulation of artificial intelligence, currently concentrated in Bill No. 2,338 of 2023, approved by the Federal Senate on 10 December 2024 and still pending deliberation in the Chamber of Deputies, whose rapporteur in the special committee has not yet delivered an opinion. The bill adopts an architecture of classification by degree of risk and secures to the person affected by a high-risk system, in Article 6, the right to an explanation of the decision, the right to contest and to request review, and the right to human review. The high-risk list in Article 14, comprising twelve headings, does not however mention credit assessment, unlike Annex III, point 5(b), of Regulation (EU) 2024/1689. The closest heading is item IV, on the assessment of criteria of access, eligibility, granting, review, reduction or revocation of private and public services regarded as essential. The classification of algorithmic credit will therefore depend on secondary regulation: Articles 15 and 16 confer on the governance system created by the bill and on the sectoral authorities the power to identify new high-risk headings and to establish, with precision, the list applicable to the regulated market. That delegation reinforces the point made in this section, because it transfers to the Central Bank the task of stating whether, and on what conditions, credit decision-making models are to be treated as high risk. Until the legislative process is concluded, however, the parameters proposed here must rest on the law in force, and not on projected legislation.

One detail of the perimeter deserves attention, and it is worth stating precisely. Article 1(1) of CMN Resolution No. 4,949/2021 excludes from its reach, among other entities, payment institutions, which are to observe the regulation issued by the Central Bank in the exercise of its statutory powers. The exclusion does not, however, produce a regulatory vacuum. BCB Resolution No. 155/2021 establishes for those institutions, and for consortium administrators, similar principles and procedures of conduct governing dealings with customers and users. And credit fintechs constituted as a Direct Credit Company or as a Peer-to-Peer Lending Company are financial institutions, a qualification conferred on them by Articles 3 and 7 of CMN Resolution No. 4,656 of 26 April 2018, repealed with effect from 1 January 2025 by Article 37(IV) of CMN Resolution No. 5,177 of 26 September 2024. The repeal does not displace the classification. CMN Resolution No. 4,949/2021 reaches financial institutions and the other institutions authorised to operate by the Banco Central do Brasil, and none of these companies appears in its list of exclusions in Article 1(1). They therefore remain fully subject to it. Three situations should not be confused: payment institutions, which do not extend credit on their own account and are governed by a mirrored regime; financial institutions specialising in digital credit, covered by the general conduct regime; and partnership or conglomerate arrangements, in which digital origination rests on a licensed financial institution, on which the duty falls. The perimeter problem that remains is a different and narrower one. It reaches, on one side, providers operating outside the authorised perimeter and, on the other, the absence of specific supervisory expectations regarding automated decision-making models, which none of the regimes cited sets out. It is this gap, and not the formal exclusion of payment institutions, that section 5.2 takes up again.

Amparo and Prado (2024) reach a convergent diagnosis: Brazilian legal regulation of discrimination in access to credit is scattered, spread across various instruments in the fields of consumer law, financial regulation, international law and anti-discrimination legislation. Motta and Rosa (2022, p. 151) conclude that the regulatory regime of Open Finance is not capable of supplying, on its own, nor is that its purpose, all the solutions to the legal questions that will arise in a data-driven financial system, illustrating the point with the potential conflict between non-transparent automated decisions and algorithmic processing methods supposedly protected by trade secret. The authors anticipate that resolving such conflicts will require closer regulatory engagement by the Central Bank and the National Monetary

Council with the effects of artificial intelligence on the functioning of the financial system, and close cooperation between those regulators and the National Data Protection Authority (Motta; Rosa, 2022, p. 151).

Fragmentation produces three undesirable effects. The first is uncertainty for the regulated entity, which must reconcile transparency duties arising from distinct regimes, with non-coinciding standards of sufficiency. The second is enforcement gaps: no authority holds, on its own, the competence and the technical capacity to audit the data, the model and the conduct. The third, and the most serious for the consumer, is the dispersal of the burden. It falls to the affected person to identify which regime to invoke and before which authority, which reintroduces by procedural means the asymmetry that the substantive rules were meant to correct.

The appropriate response is not necessarily a new statute, but institutional coordination among the existing regimes: formalised cooperation between supervisors, common minimum standards of documentation and explanation, and the designation of primary competence for the audit of credit models. To this must be added the articulation, by the Central Bank and the National Monetary Council, of expectations as to the governance of decision-making models that affect customers.

5.2 International references and regulatory trends

The international references are mobilised in a functional key: the aim is not to reconstruct the legislative history of each jurisdiction, but to extract regulatory functions that can be reproduced within the Brazilian arrangement. The criteria governing the selection were set out in the Introduction.

From the European Union, the function extracted is that of risk qualification tied to substantive obligations. Regulation (EU) 2024/1689 classifies creditworthiness assessment and credit scoring as a high-risk use, subjecting deployers to duties of use in accordance with the provider's instructions, appropriate human oversight, monitoring and record-keeping, under Article 26, to the prior fundamental rights impact assessment of Article 27(1), and to the duty of explanation of Article 86(1). The postponement introduced by Regulation (EU) 2026/1744 does not alter the regulatory function extracted here, but shifts its temporal marker: the duties under Chapter III become enforceable, for these systems, from 2 December 2027.

From the United Kingdom, the function extracted is that of disciplining the model life cycle, organised by Supervisory Statement SS1/23 of the Prudential Regulation Authority, in its consolidated version of April 2026. A limit of scope should be recorded: the document is addressed to banks, building societies and PRA-designated investment firms that hold permission to use internal models in the calculation of regulatory capital requirements, and does not reach institutions without such permission, nor branches of third-country firms operating in the United Kingdom. Even so, the tiering structure by materiality and complexity, with explicit factors for alternative data and for the potential for bias, is transposable.

From the United States, what is extracted is above all a lesson in caution. The recent trajectory shows that institutional arrangements built upon non-binding supervisory guidance are reversible at speed. Circular 2022-03 of the Consumer Financial Protection Bureau, issued on 26 May 2022, maintained that the Equal Credit Opportunity Act and Regulation B do not permit lenders to use complex algorithms where doing so prevents them from providing the specific and accurate reasons for an adverse action. It was formally withdrawn on 12 May 2025, in an en bloc withdrawal of sixty-seven guidance documents, together with Circular 2023-03. Its invocation as guidance in force is today incorrect, although it remains useful for reconstructing the debate. Add to this the replacement of SR 11-7 by SR 26-2 in April 2026, with generative artificial intelligence excluded from its scope. The functional lesson is clear: institutional parameters that aspire to endure must rest on rules of appropriate hierarchical rank, and not on revocable supervisory guidance.

CC BY NC SA

From the multilateral plane, the function extracted is that of proportionality. Crisanto et al. (2024, p. 2) identify six areas in which specific guidance may be necessary: the governance framework, with clear allocation of roles across the life cycle and specification of the role of human intervention; expertise and skills; model risk management, including as regards explainability; data governance and management; new participants and business models; and the regulatory perimeter applicable to third parties. They further record that the risks of artificial intelligence in financial institutions are, to a large extent, the same ones with which authorities have traditionally been concerned, even if some may intensify.

That finding has a strategic consequence for Brazil. If the risks are for the most part known, the primary regulatory response need not create an entirely new regime. It must make explicit how the existing regimes apply to decisions mediated by artificial intelligence, with specific regulation only where the technology introduces a new problem. The two most evident points are explainability and meaningful human oversight. These are the points on which Crisanto et al. (2024, p. 25) recommend action: financial authorities may, as a first step, define basic concepts and provide guidance on the qualities to be considered in selecting explainability techniques and in assessing their effectiveness, as well as update consumer protection regulation to articulate what types of disclosure are required when individuals are refused credit on the basis of artificial intelligence decisions.

There remains the perimeter concern, acute in Brazil. Crisanto et al. (2024, p. 27) observe that securing compliance with regulatory expectations by non-bank lenders is a specific challenge, above all as regards new entrants that use digital channels and rely on alternative data, and that non-bank lenders with digital business models are more established users of artificial intelligence. Given the significance of payment institutions and credit fintechs in the Brazilian market, coherence of the perimeter is, in our setting, a condition of effectiveness for any of the parameters proposed. As seen in section 5.1, the difficulty does not lie in the personal scope of the conduct regimes, today reasonably complete as between financial institutions and payment institutions. It lies in the activity of providers situated outside the authorised perimeter and, above all, in the absence of supervisory expectations addressed specifically to the decision-making models that determine access to credit.

Conclusion

Digitalisation has expanded the capacity for risk assessment and may reduce informational barriers to access to credit. The empirical evidence is consistent. Alternative data and machine learning make assessable profiles previously illegible to traditional models, with discriminatory power comparable to that obtained for customers with an established history (Berg et al., 2020, p. 2882). In a country where credit penetration, when measured in 2018, stood at 11% in the income bracket up to one minimum wage, owing to difficulty in evidencing income and collateral (Banco Central do Brasil, 2018, p. 26), the promise is not rhetorical.

But the same system can convert social inequalities into mathematical correlations and reproduce them under the appearance of technical neutrality. Barocas and Selbst (2016, p. 692) show that better data and more attributes tend to bring the model closer to the exact extent of existing inequality. Fuster et al. (2022, p. 5) record that the predictive gains of machine learning are not distributed equally across groups. Technical improvement, considered in isolation, does not guarantee distributive improvement.

Hence the article's final thesis. Financial citizenship in markets increasingly mediated by algorithms is not measured by the quantitative expansion of access to credit. It depends on the institutional quality of the decisions that define who obtains credit, on what terms and with what possibilities of understanding, review and contestation. The formulation is not external to positive law. It follows from the definition of financial citizenship adopted by the Banco Central do Brasil, which incorporates consumer protection, dispute resolution mechanisms and equitable treatment without discrimination as constitutive conditions of the concept (Banco Central do Brasil, 2018, p. 8, 51).

CC BY NC SA

The analytical shift proposed is to reposition effective contestability, and not explainability, as the central legal criterion of the legitimacy of algorithmic credit. Explaining is a necessary condition. Contesting is the condition that gives explaining its point. A right to explanation unaccompanied by powers of correction, of submission of additional elements and of substantive review produces ritualistic transparency: the duty to inform is discharged without any change in the legal position of the affected person. The doctrine of contestability by design supplies the vocabulary. A human in the loop is insufficient if there is no possibility of dialectical exchange between the subject of the decision and the human controller (Alfrink et al., 2023, p. 615).

The Brazilian normative material for this construction already exists, albeit dispersed: the right to review under Article 20 of Law No. 13,709/2018, with the audit for discriminatory aspects under Article 20(2); the right to know the principal elements and criteria under Article 5(IV) of Law No. 12,414/2011, with the prohibition of excessive and sensitive information under Article 3(3); the duties of objectivity, clarity and correction under Article 43 of the Consumer Protection Code; the principles of data quality and non-discriminatory treatment in the Open Finance regulation; and the duty of fair and equitable treatment under Article 3(II) of CMN Resolution No. 4,949/2021. The case law of the Superior Court of Justice, in enshrining in Súmula No. 550 the consumer's right to clarification as to the personal information assessed and the sources of the data considered, already recognises the core of the right. What is missing is institutional coordination and the articulation of supervisory expectations.

The regulatory challenge does not consist in choosing between innovation and protection. It consists in building institutions capable of making innovation financially inclusive and legally legitimate. On that view, data governance and model governance, impact assessment, proportionate transparency, contestability and accountability are not obstacles external to credit innovation. They are conditions for the expansion of digital credit to be recognised, properly, as an expansion of financial citizenship.

References

ALFRINK, Kars; KELLER, Ianus; KORTUEM, Gerd; DOORN, Neelke. **Contestable AI by design**: towards a framework. **Minds and Machines**, v. 33, n. 4, p. 613-639, 2023. Disponível em: https://doi.org/10.1007/s11023-022-09611-z. Acesso em: 30 ago. 2026.

ALMADA, Marco. Human intervention in automated decision-making: toward the construction of contestable systems. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 17., 2019, Montreal. **Proceedings** [...]. New York: Association for Computing Machinery, 2019. p. 2-11. Disponível em: https://doi.org/10.1145/3322640.3326699. Acesso em: 30 ago. 2026.

AMPARO, Thiago; PRADO, Viviane Muller. **Racismo creditício no Brasil e nos EUA**: risco discriminatório no acesso a crédito. **Revista Direito GV**, São Paulo, v. 20, e2422, 2024. Disponível em: https://doi.org/10.1590/2317-6172202422. Acesso em: 30 ago. 2026.

BANCO CENTRAL DO BRASIL. **Relatório de Cidadania Financeira 2018.** Brasília: Banco Central do Brasil, 2018. Disponível em: https://www.bcb.gov.br/nor/relcidfin/docs/relatorio_cidadania_financeira.pdf. Acesso em: 30 ago. 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/RCF/Relatorio_de_Cidadania_Financeira_2021.pdf. Acesso em: 28 set. 2026.

CC BY NC SA

BANCO CENTRAL DO BRASIL. **Relatório de Cidadania Financeira 2025.** Brasília: Banco Central do Brasil, 2025. Disponível em: https://www.bcb.gov.br/content/cidadaniafinanceira/documentos_cidadania/RIF/relatorio_de_cidadania_financeira_2025.pdf. Acesso em: 28 set. 2026.

BANCO CENTRAL DO BRASIL. **Resolução BCB 155, de 14 de outubro de 2021.** Dispõe sobre princípios e procedimentos a serem adotados no relacionamento com clientes e usuários de produtos e de serviços pelas administradoras de consórcio e pelas instituições de pagamento autorizadas a funcionar pelo Banco Central do Brasil. Brasília: BCB, 2021. Disponível em: https://www.bcb.gov.br/estabilidadefinanceira/buscanormas. Acesso em: 28 set. 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://doi.org/10.15779/Z38BG31. Acesso em: 30 ago. 2026.

BARTLETT, Robert; MORSE, Adair; STANTON, Richard; WALLACE, Nancy. **Consumer-lending discrimination in the FinTech era**. **Journal of Financial Economics**, v. 143, n. 1, p. 30-56, 2022. Disponível em: https://doi.org/10.1016/j.jfineco.2021.05.047. Acesso em: 30 ago. 2026.

BERG, Tobias; BURG, Valentin; GOMBOVIĆ, Ana; PURI, Manju. **On the rise of FinTechs**: credit scoring using digital footprints. **The Review of Financial Studies**, v. 33, n. 7, p. 2845-2897, 2020. DOI: 10.1093/rfs/hhz099. Versão preliminar: Cambridge, MA: National Bureau of Economic Research, jul. 2018 (NBER Working Paper, n. 24551). Disponível em: https://www.nber.org/papers/w24551. Acesso em: 28 set. 2026.

BRASIL. **Lei 8.078, de 11 de setembro de 1990.** Dispõe sobre a proteção do consumidor e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, 12 set. 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 30 ago. 2026.

BRASIL. **Lei 12.414, de 9 de junho de 2011.** Disciplina a formação e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para formação de histórico de crédito. Diário Oficial da União: seção 1, Brasília, DF, 10 jun. 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12414.htm. Acesso em: 30 ago. 2026.

BRASIL. **Lei 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 30 ago. 2026.

BRASIL. Congresso Nacional. **Projeto de Lei 2.338, de 2023.** Dispõe sobre o uso da inteligência artificial. Aprovado pelo Senado Federal em 10 dez. 2024; em tramitação na Câmara dos Deputados. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 28 set. 2026.

BRASIL. Superior Tribunal de Justiça. **Súmula 550.** Segunda Seção, julgado em 14 out. 2015. Diário da Justiça Eletrônico, Brasília, DF, 19 out. 2015. Disponível em: https://scon.stj.jus.br/SCON/sumstj/. Acesso em: 28 set. 2026.

BRASIL. Superior Tribunal de Justiça (Segunda Seção). **Recurso Especial 1.419.697/RS (Tema 710).** Relator: Ministro Paulo de Tarso Sanseverino, julgado em 12 nov. 2014, DJe 17 nov. 2014. Disponível em: https://processo.stj.jus.br/repetitivos/temas_repetitivos/. Acesso em: 28 set. 2026.

CHENEY-LIPPOLD, John. **Jus algoritmi**: how the National Security Agency remade citizenship. **International Journal of Communication**, v. 10, p. 1721-1742, 2016. Disponível em: https://ijoc.org/index.php/ijoc/article/download/4480/1618. Acesso em: 30 ago. 2026.

CONSELHO MONETÁRIO NACIONAL. **Resolução CMN 4.557, de 23 de fevereiro de 2017.** Dispõe sobre a estrutura de gerenciamento de riscos, a estrutura de gerenciamento de capital e a política de divulgação de informações. Brasília: CMN, 2017. Disponível em: https://normativos.bcb.gov.br/Lists/Normativos/Attachments/50344/Res_4557_v4_P.pdf. Acesso em: 28 set. 2026.

CONSELHO MONETÁRIO NACIONAL. **Resolução CMN 4.656, de 26 de abril de 2018.** Dispõe sobre a sociedade de crédito direto e a sociedade de empréstimo entre pessoas. Brasília: CMN, 2018. Disponível em: https://www.bcb.gov.br/estabilidadefinanceira/buscanormas. Acesso em: 28 set. 2026.

CONSELHO MONETÁRIO NACIONAL. **Resolução CMN 4.949, de 30 de setembro de 2021.** Dispõe sobre princípios e procedimentos a serem adotados no relacionamento com clientes e usuários de produtos e de serviços. Brasília: CMN, 2021. Disponível em: https://www.in.gov.br/web/dou/-/resolucao-cmn-n-4.949-de-30-de-setembro-de-2021-350015767. Acesso em: 28 set. 2026.

CONSELHO MONETÁRIO NACIONAL. **Resolução CMN 5.177, de 26 de setembro de 2024.** Dispõe sobre a política de remuneração de administradores das instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil. Brasília: CMN, 2024. Disponível em: https://www.bcb.gov.br/estabilidadefinanceira/buscanormas. Acesso em: 29 set. 2026.

CONSELHO MONETÁRIO NACIONAL; BANCO CENTRAL DO BRASIL. **Resolução Conjunta 1, de 4 de maio de 2020.** Dispõe sobre a implementação do Open Finance, na redação dada pela Resolução Conjunta 4, de 24 de março de 2022. Brasília: CMN/BCB, 2020. Disponível em: https://normativos.bcb.gov.br/Lists/Normativos/Attachments/51028/Res_Conj_0001_v7_L.pdf. Acesso em: 28 set. 2026.

CONSUMER FINANCIAL PROTECTION BUREAU. **Consumer Financial Protection Circular 2022-03**: adverse action notification requirements in connection with credit decisions based on complex algorithms. Washington, DC, 26 maio 2022. 87 Fed. Reg. 35864, 14 jun. 2022. Retirada em 12 maio 2025 (90 Fed. Reg. 20084). Disponível em: https://www.federalregister.gov/documents/2025/05/12/2025-08286/interpretive-rules-policy-statements-and-advisory-opinions-withdrawal. Acesso em: 28 set. 2026.

CRISANTO, Juan Carlos; LEUTERIO, Cris Benson; PRENIO, Jermy; YONG, Jeffery. **Regulating AI in the financial sector**: recent developments and main challenges. Basel: Bank for International Settlements, dez. 2024. (FSI Insights on Policy Implementation, n. 63). Disponível em: https://www.bis.org/fsi/publ/insights63.pdf. Acesso em: 30 ago. 2026.

ESTADOS UNIDOS. Board of Governors of the Federal Reserve System; Office of the Comptroller of the Currency; Federal Deposit Insurance Corporation. **Supervisory Letter SR 26-2**: revised guidance on model risk management. Washington, DC, 17 abr. 2026. Disponível em: https://www.federalreserve.gov/supervisionreg/srletters/SR2602.htm. Acesso em: 30 ago. 2026.

FOURCADE, Marion; HEALY, Kieran. **Classification situations**: life-chances in the neoliberal era. **Accounting, Organizations and Society**, v. 38, n. 8, p. 559-572, 2013. Disponível em: https://doi.org/10.1016/j.aos.2013.11.002. Acesso em: 30 ago. 2026.

CC BY NC SA

FUSTER, Andreas; GOLDSMITH-PINKHAM, Paul; RAMADORAI, Tarun; WALTHER, Ansgar. **Predictably unequal?** The effects of machine learning on credit markets. **The Journal of Finance**, v. 77, n. 1, p. 5-47, 2022. Disponível em: https://doi.org/10.1111/jofi.13090. Acesso em: 30 ago. 2026.

MOTTA, Bernardo Rocha da; ROSA, Marcus Paulus de Oliveira. Open Banking, Big Data e Inteligência Artificial: como tudo está conectado na regulação de um sistema financeiro e de pagamentos movido a dados? **Revista da Procuradoria-Geral do Banco Central**, Brasília, v. 16, n. 1, p. 132-154, jun. 2022. Disponível em: https://revistapgbc.bcb.gov.br/revista/article/download/1163/83. Acesso em: 30 ago. 2026.

REINO UNIDO. Prudential Regulation Authority. **Supervisory Statement SS1/23**: model risk management principles for banks. London: Bank of England, maio 2023; versão consolidada de 23 abr. 2026. Disponível em: https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss. Acesso em: 28 set. 2026.

STIGLITZ, Joseph E.; WEISS, Andrew. **Credit rationing in markets with imperfect information**. **The American Economic Review**, v. 71, n. 3, p. 393-410, jun. 1981. Disponível em: https://www.jstor.org/stable/1802787. Acesso em: 28 set. 2026.

TOMASELLO, Federico. **Digital civics and algorithmic citizenship in a global scenario**. **Philosophy & Technology**, v. 36, art. 39, p. 1-5, 2023. Disponível em: https://doi.org/10.1007/s13347-023-00638-3. Acesso em: 30 ago. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024**, que estabelece regras harmonizadas em matéria de inteligência artificial. Jornal Oficial da União Europeia, L, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Acesso em: 30 ago. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2026/1744 do Parlamento Europeu e do Conselho, de 8 de julho de 2026**, que altera os Regulamentos (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 no que respeita à simplificação da aplicação de regras harmonizadas em matéria de inteligência artificial. Jornal Oficial da União Europeia, L, 24 jul. 2026. Disponível em: https://eur-lex.europa.eu/eli/reg/2026/1744/oj. Acesso em: 28 set. 2026.