香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

人工智能與法律研究中心
Center for Artificial Intelligence for Law

AI GOVERNANCE / CONSUMER PROTECTION

# Consumer Protection in AI–Governed Credit Markets

## Algorithmic Bias and the Right to Explanation



**Yidan Huang** · **Sirui Han***

University of Illinois Journal of Law, Technology & Policy
2026 · Volume 2026 · Issue 2 · pp. 37–115

# CONSUMER PROTECTION IN AI–GOVERNED CREDIT MARKETS: ALGORITHMIC BIAS AND THE RIGHT TO EXPLANATION

*Yidan Huang*[†]
*Sirui Han*[‡]

*Focusing on the European Union and China, and using the United States as a comparative reference point, this article examines whether and to what extent the legal frameworks of the European Union and China recognize and enforce a right to algorithmic explanation in automated credit scoring as a form of artificial intelligence–enabled consumer protection. Automated credit scoring has become a core decision infrastructure in contemporary credit markets, influencing access to credit and consumers' economic opportunities; however, it often relies on opaque models and large-scale personal data, raising concerns regarding bias and the lack of clear reasons for adverse outcomes. This article conducts a comparative doctrinal analysis to demonstrate that bias in credit scoring constitutes a legally cognizable harm that is difficult to contest where decision-making is insulated by evidential asymmetry, fragmented legal duties, and assertions of commercial secrecy. Furthermore, it clarifies the situations when credit scoring should be treated as solely automated, focusing on systems with minimal human involvement or where scores effectively predetermine outcomes. The article explains why contractual, tort-*

† Ms. Yidan Huang is a doctoral student at the Hong Kong University of Science and Technology, researching AI governance, legal technology, and consumer protection. Her work has been published in *Artificial Intelligence and Law* and presented at CHI 2026, ICA 2026, and the 2026 Global Conference on Empirical Legal Studies. She previously worked as a Data Analyst at the Hong Kong Generative AI Research and Development Center, contributing to legal technology projects through data analysis and experimental evaluation. She holds a Juris Doctor from the City University of Hong Kong.

‡ Prof. Sirui Han is an Assistant Professor at HKUST, Director of its Center for Artificial Intelligence for Law, and Associate Director of the HKUST-DXM AI for Finance Joint Laboratory. Her research spans law and economics, large-language-model reasoning and alignment, and legal and financial AI, with publications in *Artificial Intelligence and Law* and the *International Review of Law and Economics*, alongside papers at NeurIPS, ICLR, AAAI, ACL, EMNLP, and CHI; she is also the lead author of the *Hong Kong Generative Artificial Intelligence Technical and Application Guideline* and Principle Investor of developing the LexiHK legal AI assistant for HKSAR Department of Justice, which received a bronze medal at the 2025 International Exhibition of Inventions of Geneva.

Correspondence regarding this work should be addressed to Prof. Sirui Han, the corresponding author, at the Hong Kong University of Science and Technology (email: siruihan@ust.hk).

*based, and information-based remedies frequently fail to provide timely and meaningful redress. Based on the analysis, it proposes a structured right to algorithmic explanation that is tailored to credit markets, specifying duty bearers, trigger conditions, and procedural mechanisms for review and challenge. The article concludes that both jurisdictions confront the shared challenge of translating formal obligations toward transparency into effective consumer contestability in algorithmically governed credit markets.*

Keywords: algorithmic explanation; credit scoring; consumer protection; algorithmic bias; comparative law

TABLE OF CONTENTS

I. INTRODUCTION....39
II. AUTOMATED CREDIT SCORING....49
A. Automated decision-making....52
B. Hybrid legal interest....57
III. REGULATING CREDIT SCORING AND LEGAL EFFECTS....60
A. The contractual framework....61
B. Evidentiary challenges in tort remedies....65
IV. EXPLAINABILITY, RESPONSIBILITY, AND TRADE SECRECY....70
A. Are algorithms explainable?....71
B. Responsibility in automated credit governance....77
C. Do explanation rights conflict with trade secrets?....80
V. FROM NOTICE TO CONTESTABILITY....82
A. Is the right of access or to be informed adequate?....82
B. Is the right to explanation clearly recognized?....85
C. The U.S. Adverse-Action Model of Credit Explanation....89
VI. LEGAL CONSTRUCTION OF THE RIGHT TO EXPLANATION....94
A. The right holder and duty bearer....95
B. Modes of explanation....99
C. Gradient requirements....102
1. Proportionality....103
2. Trigger thresholds....103
3. The content gradient....105
4. Trade secrets and third-party data....107
D. Procedural design for disclosure....109
VII. CONCLUSION AND FUTURE WORK....112

## I. INTRODUCTION

Personal credit scoring, as a sociotechnical artifact, illustrates a historical progression from interpersonal reputation to algorithmic inference.[1] Early forms of credit evaluation, such as the role of *paojie (跑街)*[2] in late Qing China or manual credit reporting in 19th-century America[3], were based on local knowledge, interpersonal observation, and moral assessment.[4] Over time, longstanding credit-evaluation practices have evolved into digitized scoring systems built on large-scale data collection, statistical modeling, and increasingly machine-learning techniques.[5] Rather than replacing subjective judgment with neutral computation, the shift relocated credit allocation from interpersonal assessment to data-driven evidentiary construction.[6] Traditional credit scoring relied heavily on structured financial attributes, including formal identification, repayment history, bank transactions, income records, asset values, and other credit-file information.[7] These inputs often worked poorly for underbanked or thin-file borrowers, whose creditworthiness was not legible through conventional records.[8] Machine-learning systems respond to that constraint by extracting latent credit signals from fragmented behavioral data, alternative data, and large-scale correlations.[9] More recent AI and LLM-based systems extend this trend by transforming unstructured information, such as loan applications, customer communications, bank statements, business descriptions, compliance documents, and analyst reports, into decision-relevant signals.[10] In this sense, AI-

1. JOSH LAUER, CREDITWORTHY: A HISTORY OF CONSUMER SURVEILLANCE AND FINANCIAL IDENTITY IN AMERICA 211 (Colum. Univ. Press, 2017).

2. *See* JIANGUO SUN, THE EVOLUTION OF CREDIT: A STUDY OF THE CHINA CREDIT BUREAU IN SHANGHAI, 74–83 (China Soc. Sci. Press, 2007) ("[m]erchant houses largely relied upon intermediaries such as street runners, who circulated within networks of acquaintances to ascertain debtors' character, performance of obligations, and social standing. Credit did not stem from institutionalised records, but was founded upon interpersonal relationships and moral assessment. . .").

3. Claire Brennecke, *Information Acquisition in Antebellum U.S. Credit Markets: Evidence from Nineteenth Century Credit Reports*, at 7–9 (FDIC Ctr. Fin. Rsch., Working Paper 2016-04, 2016) (discussing early manual credit reporting methods in 19th century America).

4. LAUER, *supra* note 1, at 27; SUN, *supra* note 2, at 74.

5. LAUER, *supra* note 1, at 183.

6. *Id.*

7. *See* David Mhlanga, *Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment*, 9 INT'L J. FIN. STUD., at 12 (2021) (highlighting how explainable AI (XAI) models can overcome traditional information asymmetry by generating transparent credit risk assessments from alternative data, thereby expanding financial inclusion to previously unbanked populations; "The traditional data that was used in generating credit scores include formal identification, bank transactions, credit history, income statements and asset value. In emerging markets, the underbanked individuals. . . find it difficult to access the traditional forms of collateral or identification. . .").

8. *Id.*

9. *See* Tobias Berg et al., *On the Rise of FinTechs—Credit Scoring Using Digital Footprints*, at 21 (Nat'l Bureau Econ. Rsch., Working Paper No. 24551, 2018) (showing that digital-footprint variables can complement credit-bureau information and affect access to credit); Amir E. Khandani et al., *Consumer Credit-Risk Models via Machine-Learning Algorithms*, 34 J. BANKING & FIN. 2767, 2767–69 (2010) (showing that machine-learning models can improve consumer credit-risk classification).

10. *See* Yinheng Li et al., *Large Language Models in Finance: A Survey*, PROC. FOURTH ACM INT'L CONF. AI FIN., Nov. 2023, at 375 (explaining that LLMs can transform unstructured financial text into decision-relevant information); *see also* Allen H. Huang & Haifeng You, *Artificial Intelligence in Financial Decision Making*, HANDBOOK FIN. DECISION MAKING, Oct. 2022, at 22 (discussing how AI extracts information from

powered credit scoring may improve processing productivity, reduce underwriting costs, identify borrowers overlooked by conventional scorecards, and support more scalable financial decision-making.[11]

At the same time, the technical development expands the legal and institutional stakes of credit scoring.[12] Stronger prediction does not guarantee fairer credit allocation.[13] In fact, machine-learning credit models can improve prediction while still distributing credit-market gains unevenly across borrower groups.[14] Digital lending can reduce some forms of human discretion yet preserve discriminatory outcomes through pricing, proxy variables, and model-mediated underwriting.[15] Big-data credit scoring may also infer creditworthiness from shopping patterns, social networks, device behavior, location traces, or other variables that are difficult for consumers to observe, understand, or correct.[16] Thus, contemporary models can dynamically learn from new correlations and draw on alternative data sources, widening predictive reach and governance stakes.[17] These developments have repositioned credit scoring from a narrow instrument of risk assessment to a hybrid infrastructure operating across market

structured and unstructured data and complements human financial decision-making); *see* Junyu Luo et al., *FINMME: Benchmark Dataset for Financial Multi-Modal Reasoning Evaluation*, 1 PROC. 63RD ANN. MEETING ASS'N FOR COMPUT. LINGUISTICS, 29465, 29465–66 (2025) (finding that leading multimodal models still underperform on knowledge-intensive financial tasks and that financial evaluation should incorporate hallucination penalties because the sector has low tolerance for inaccuracies); *see* Zhizhuo Kou et al., *Automate Strategy Finding with LLM in Quant Investment*, ASS'N FOR COMPUT. LINGUISTICS, NOV. 2025, at 18517–20 (the study illustrates how LLM-based financial systems can convert financial research, numerical data, market charts, and other multimodal materials into decision-relevant trading signals).

11. *See* Ziwei Yi et al., *Artificial Intelligence in Accounting and Finance: Challenges and Opportunities*, 11 *IEEE ACCESS, at* 1–4 (2023) (surveying AI applications in accounting and finance); *see* Xiao-Lin Zheng et al., *FinBrain: When Finance Meets AI 2.0**, 20 FRONTIERS INFO. TECH. & ELEC. ENG'G 914, 915 (2019) (explaining ways in which AI will increase efficiency in financial markets and reduce waste); *see* Shuaishuai Gong et al., *Artificial Intelligence in Financial Markets: A Literature Review and Research Agenda*, 28 CHINA ACCT. & FIN. REV., at 1–2, 15 (2026) (reviewing AI applications in finance and identifying algorithmic trading, credit scoring, fraud detection and regulatory compliance as major areas of development; noting that AI can "enhance predictive accuracy. . . and expand credit access," while also raising challenges of "model interpretability, fairness and systemic risk.").

12. *See* Andreas Fuster et al., *Predictably Unequal? The Effects of Machine Learning on Credit Markets*, 77, J. FIN. 5, 8 (2022) (explaining the default propensities across racial and ethnic groups in AI-LLM based credit models); *See* Ahmed Mohsen, *How responsibly deploying AI credit scoring models can progress financial inclusion*, WORLD ECON. F. (Oct. 1, 2025), https://www.weforum.org/stories/financial-and-monetary-systems/how-responsibly-deploying-ai-credit-scoring-models-can-progress-financial-inclusion/ [https://perma.cc/Y4G8-SQKS] (describing some of the potential benefits of AI integration in credit scoring, as well as some of risks of irresponsible use of AI in this space).

13. Fuster et al., *supra* note 12, at 6; Mohsen, *supra* note 12.

14. Fuster et al., *supra* note 12, at 5–6; Mohsen, *supra* note 12.

15. .*See* Fuster et al., *supra* note 12, at 8 (finding that machine learning can improve prediction while unevenly distributing benefits across borrower groups); Robert Bartlett et al., *Consumer-Lending Discrimination in the FinTech Era,* 143 J. FIN. ECON. 30, 55 (2022) (finding that discrimination persists in fintech-era lending).

16. *See* Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J.L. & TECH. 148, 182, 186 (2016) (explaining that big-data credit scoring may rely on variables such as shopping behavior, social networks, and other nontraditional signals that consumers may struggle to understand or challenge).

17. *See* Hesam Kamyab et al., *The Latest Innovative Avenues for the Utilization of Artificial Intelligence and Big Data Analytics in Water Resource Management*, 20 RESULTS ENG'G, at 4 (2023) (discussing how enhancing the interpretability of complex AI models is essential for ensuring decision-makers can trust and understand the reasoning behind AI-driven recommendations in water management; "The iterative nature of AI enables continuous strategy refinement as new data become available. . . .").

governance and regulatory regimes.[18] Simultaneously, they have intensified proprietary claims over model logic and consolidated informational power among those who control data and computations, thereby creating a technology-enabled accountability enclave.[19]

In artificial intelligence (AI)-governed credit markets, automated credit scoring is normatively and legally difficult not only due to the complexity of the scoring logic[20] but also because its operational design systematically deprives consumers of the necessary conditions for meaningful contestability.[21] Even when a credit product enumerates broad scoring dimensions, such as account tenure, identity completeness, payment activity, and scenario diversity, such disclosures do not enable verification,[22] because one source of opacity lies in the absence of information about the mapping of each dimension from data to score, the relative contribution of each factor, and the points at which errors can enter the scoring system.[23] Thus, the resulting asymmetry is not an accidental shortcoming;[24] rather, it is a predictable feature of a market environment that treats model logic as proprietary and often frames such data as confidential business information,[25] thereby narrowing the space for independent audits and transforming secrecy into a structural substitute for justification.[26] In such scenarios, the consumer cannot perform the minimal legal tasks underpinning fairness in

18. *See generally* Fuster et al., *supra* note 12 (describing the effects of machine learning on credit markets and on credit scoring).

19. *See* Xinbao Zhang (张新宝), *Gè Rén Xìn Xī Shōu Jí: Gào Zhī Tóng Yì Yuán Zé Shì Yòng De Xiàn* Zhì (个人信息收集：告知同意原则适用的限制) [*Collection of Personal Information: Restricting the Application of the Principle of Informed Consent*], 33 BǏ JIÀO FǍ XUÉ YÁN JIŪ (比较法学研究) [J. COMP. L.] 111 (2019) (arguing that as the principle of "informed consent" is restricted for large-scale data collection, the need for algorithmic explainability becomes paramount to justify automated decisions and maintain public trust); *see* Sirui Han et al., *Revolutionizing Finance with Bitcoin and Blockchain: A Literature Review and Research Agenda*, 26 CHINA ACCT. & FIN. REV. 413, 426–27 (2024) (noting that Bitcoin regulation includes "consumer financial protection" and should advance understanding of the "interaction among the disciplines of finance, economics, law and technology.").

20. Łukasz Górski, *Against Explainable Artificial Intelligence in Law: Why Justifiable AI Matters. A Credit Scoring Example*, INTERDISC. CTR. MATHEMATICAL & COMPUTATIONAL MODELING, Aug. 2026, at 2, 6.

21. *Id.* at 16–17.

22. *See* Hurley & Adebayo, *supra* note 16, at 166 (explaining that going beyond traditional credit bureau data still does not enable verification); *see* Korin Munsterman, *When Algorithms Judge Your Credit: Understanding AI Bias in Lending Decisions*, ACCESSIBLE L., Spring 2025, https://www.accessiblelaw.untdallas.edu/post/when-algorithms-judge-your-credit-understanding-ai-bias-in-lending-decisions [https://perma.cc/S2WK-JPWY] ("The 'black box' nature of these algorithms compounds these problems. When credit decisions move from straightforward formulas to complex machine learning models, decisions become increasingly opaque, making it nearly impossible for consumers—or even regulators—to understand why a particular decision was made.").

23. *See e.g.*, *Wēi Xìn Zhī Fù Fēn Yòng Hù Fú Wù Xié Yì* (微信支付分用户服务协议) [*Wechat Pay Score User Service Agreement (Version 20241121)*], TÉNG XÙN (腾讯) [TENCENT] (Mar. 10, 2025), https://posts.tenpay.com/posts/54ad1eea621dc621139fbaa805165f3b.html [https://perma.cc/9XV5-6UYM] [hereinafter Tencent].

24. *See* Danielle K. Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 20–21 (2014).

25. *Id.*

26. *See* Linghan Zhang (张凌寒), *Shāng Yè Zì Dòng Huà Jué Cè De Suàn Fǎ Jiě Shì Quán Yán Jiū* (商业自动化决策的算法解释权研究) [*Research on the Algorithm Interpretation Rights of Business Automation Decision-Making*], 36 FA. LV. KE. XUE. [LEGAL SCI.] 65, 68–72 (2018) [hereinafter Zhang] (arguing that establishing a legal right to algorithmic explanation is essential for protecting individual interests and ensuring fairness in automated commercial decision-making).

private ordering, such as identifying inaccurate inputs, challenging irrelevant proxies, or demonstrating differential treatment with sufficient specificity to trigger scrutiny.[27] In particular, such withholding is one source of opacity with distributive consequences, as has been extensively documented in credit scoring and adjacent algorithmic domains.[28] Comparable accountability failures in housing[29], pretrial detention[30], and sentencing[31] reinforce the same issue; namely, when decision systems are designed such that reasons are systematically unavailable, illegality becomes harder to prove while accountability becomes easier to evade.[32]

Automated prediction systems structure access to opportunity while limiting the affected person's ability to inspect or challenge the basis of classification.[33] For instance, Pasquale and Citron critically examine the societal impact of automated prediction systems, advocating for enhanced due process, robust legal oversight, and greater ethical considerations to ensure fairness and transparency.[34] In response, the central regulatory question is not whether the model is, on average, accurate but whether the legal order supplies individuals with the

27. *See* Ye Ming (叶明) & Qiu Weiqi (邱威棋), *Mì Mì Zì Zhì De Shù Zì Píng Tái Xìn Yòng Píng Fēn: Yǐn Yōu Yǔ Guī Zhì* ("秘密自治"的数字平台信用评分：隐忧与规制) [*"Secretly Autonomous" Digital Platform Credit Scoring: Hidden Dangers and Regulation*], 23 BĚI JĪNG GŌNG YÈ DÀ XUÉ XUÉ BÀO SHÈ HUÌ KĒ XUÉ BǍN (北京工业大学学报社会科学版) [J. BEIJING UNIV. TECH. (SOC. SCI. ED.)], at 118–20 (2023) (investigating how the opacity of "secretly autonomous" credit scoring platforms creates significant risks for individuals, making regulatory demands for algorithmic explainability and transparency a necessity).

28. *See* Hurley & Adebayo, *supra* note 16, at 179 (discussing the vital need to enhance big—data credit—scoring algorithm interpretability to ensure transparency, fairness, and non-discrimination in credit assessment, as current tools face transparency issues and legal gaps; "[c]redit-scoring companies treat their data sources as proprietary trade secrets. [C]onsumers have no realistic means to understand. . . [or] challenge their scores….").

29. *See* Virginia Foggo & John Villasenor, *Algorithms, Housing Discrimination, and the New Disparate Impact Rule*, 22 COLUM. SCI. & TECH. L. REV. 1, 2 (2020) (discussing how enhancing the interpretability of algorithms in housing discrimination contexts is crucial for ensuring fairness, transparency, and compliance with the new disparate impact rule; "Depending on its interpretation in the courts, the new rule risks erecting very high barriers to future FHA plaintiffs in light of the proprietary nature of the algorithms they will be challenging. [T]he Article presents a roadmap which would allow plaintiffs to access the information necessary. . . .").

30. *See* Sandra G. Mayson, *Bias in, Bias Out*, 128 YALE L.J. 2218, 2279 (2019) (discussing how addressing algorithmic bias through interpretability and transparency is essential for ensuring fairness and accountability in automated decision-making systems; "There is no reason to think subjective risk assessment is immune. On the contrary, ample and mounting evidence has documented otherwise inexplicable racial disparities in policing, charging, pretrial detention, and sentencing."); *see also* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1098 (2019) (discussing how enhancing interpretability in algorithmic criminal justice is vital for achieving racial equity and ensuring socially acceptable outcomes; "Many of the algorithms . . . are sheltered from disclosure by trade secrets law and hence are not disclosed presently to the public. Even if they were to be disclosed in the course of litigation, it would likely be under the auspices of a protective order.").

31. *See generally* Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 842–66 (2014) (discussing how enhancing interpretability and transparency in evidence-based sentencing algorithms is crucial for preventing scientific rationalization of discrimination and ensuring fairness).

32. *See* Hurley & Adebayo, *supra* note 16, at 200 (noting the increased prevalence of big-data scoring tools which could lead to a risk the tools will be used to identify individuals susceptible to predatory loan products); *see also* Citron & Pasquale, *supra* note 24, at 15 (noting the difficulty in assessing the effectiveness of the ECOA).

33. Munsterman, *supra* note 22.

34. *See* Citron & Pasquale, *supra* note 24, at 19–20 (discussing how ensuring interpretability and accountability in automated prediction systems is vital for upholding due process and preventing unjust outcomes; "the underlying values of due process—transparency, accuracy, accountability. . . and fairness—should animate the oversight of scoring systems. . . .").

means to understand and contest adverse outcomes in an institutionally credible and practically usable manner.[35] In this context, the right to an algorithmic explanation is regarded as a key legal countermeasure.[36] However, Edwards and Veale highlight that this right currently lacks a clear legal definition, with the lack of unified standards leading to disagreements regarding its scope among different stakeholders.[37] In addition, technical difficulties and issues exist regarding attribution of responsibility.[38] Existing studies indicate that the current legal framework in the European Union (EU) exhibits significant limitations in terms of consumer protection.[39] Specifically, the Unfair Commercial Practices Directive is primarily designed for traditional marketing methods and does not offer sufficient protection against AI-driven algorithmic decision-making that harms consumer interests.[40] Though the AI Act has introduced transparency requirements,[41] its definition of "high risk" remains broad and lacks a detailed consideration of potential consumer harm in specific scenarios.[42] By contrast, the Digital Services Act focuses on platform liability but does not delve into algorithmic logic;[43] thus, its effectiveness at addressing AI-driven decision-making discrimination is limited.[44] Meanwhile, in China, financial consumers' rights are primarily protected through the application of public interest litigation under

35. *See* Gillian K. Hadfield, *The Levers of Legal Design: Institutional Determinants of the Quality of Law*, 36 J. COMP. ECON. 43, 64 (2008) (discussing how institutional design influences the interpretability and adaptability of legal systems, which is crucial for ensuring fairness and accountability in algorithmic decision-making within the legal domain; "If more information is available to a wider audience . . . an increased capacity . . . squeeze information out. . . .").

36. *See* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 209 (2019) (discussing how the General Data Protection Regulation (GDPR) mandates transparency and interpretability in algorithmic decision-making to protect individual rights and ensure accountability).

37. *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 39 (2017) (discussing how aligning algorithmic explainability with diverse stakeholder needs is essential for ensuring socially acceptable, fair, and reliable automated decision-making outputs).

38. *Id.* at 48 (questioning "where responsibility should lie" in algorithmic decision-making and warning against "confusing automation with lack of responsibility").

39. *See* Mateusz Grochowski et al, *Algorithmic Transparency and Explainability for EU Consumer Protection: Unwrapping the Regulatory Premises*, 8 CRITICAL ANALYSIS L. 43, 52 (2021) (discussing how aligning algorithmic transparency and explainability with regulatory and consumer protection goals is essential for ensuring socially acceptable and reliable outputs in automated decision-making; noting that under the Unfair Commercial Practices Directive, disclosure obligations "does not extend to information about the factors which determined a specific algorithmic outcome" and therefore "limit [their] significance for consumers.").

40. *See id.* (discussing how aligning algorithmic transparency and explainability with regulatory and consumer protection goals is essential for ensuring socially acceptable and reliable outputs in automated decision-making; noting that under the Unfair Commercial Practices Directive, disclosure obligations "do not extend to information about the factors which determined a specific algorithmic outcome" and therefore "limit [their] significance for consumers.").

41. *See* Kasia Söderlund, *High-Risk AI Transparency? On Qualified Transparency Mandates for Oversight Bodies Under the EU AI Act*, 2025 TECH. REGUL. 97, 97–98 (2025) (discussing how aligning algorithmic transparency and explainability with the qualified transparency framework under the EU AI Act is essential for ensuring socially acceptable and reliable outputs in AI systems).

42. *Id.*

43. 2022 O.J. (L 277) 17; *see* Bram Duivenvoorde, *Generative AI and the Future of Marketing: A consumer protection perspective*, 57 COMPUT. L. & SEC. REV. 1, 8–9 (2025) (noting that DSA rules "only address platform advertising," that their transparency duties have "limited" protective effect, and that discrimination safeguards are confined to narrow profiling-based prohibitions).

44. *See* Duivenvoorde, *supra* note 43 at 8–9.

the Consumer Rights and Interests Protection Law.[45] However, when facing the issues posed by automated credit scoring, notable problems arise, such as a significant information imbalance and difficulties in formulating rules due to technological complexity.[46] The lack of transparency and explanation in algorithmic decision-making further increases the difficulties faced by plaintiffs in asserting their rights.[47] Concrete platform disputes involving search ranking, algorithmic resume screening, and data driven price discrimination show that Chinese claimants already encounter comparable levels of opacity and contestability in adjacent consumer and employment settings, including refusals to explain algorithmic outputs on trade secret grounds.[48] The U.S. framework offers a partial answer, but it does not ensure that scoring systems are transparent, unbiased, or practically contestable.[49] The Fair Credit Reporting Act (FCRA) regulates consumer reports and excludes certain outdated or sensitive records, but it does not generally tell consumers what alternative data, behavioral signals, or proxy variables were used to calculate a score, how those inputs were weighted, or whether they were used to infer vulnerability rather than creditworthiness.[50] Equal Credit Opportunity Act (ECOA) and Regulation B require creditors to

45. *See generally* Junke Xu, *Who Will Protect Chinese Consumers? The Past, Present and Future of Consumer Protection Legislation in China*, 24 LOY. CONSUMER L. REV. 22 (2011) (overviewing Chinese Consumer Rights and Interests Protection Law).

46. Ye Ming (叶明) & Qiu Weiqi (邱威棋), *supra* note 27 (discussing difficulties within the Chinese credit scoring space).

47. *See* Cheng Xuejun (汪敏) & Wang Min (程雪军), *Suanfa Shehui xia Jinrong Xiaofeizhe Gongyi Susong de Fazhi Yinying* (算法社会下金融消费者公益诉讼的法治因应) [*The Rule of Law Response to Public Interest Litigation by Financial Consumers in the Algorithmic Society*], 2 FAZHI SHIWU (法治实务) [PRAC. RULE L.] 238 (2024) (discussing how enhancing the interpretability of algorithms in the context of financial consumer public interest litigation is crucial for ensuring transparent, accountable, and socially acceptable outcomes).

48. *See Jing Xiao, Wéi Shá Sōu Chū Lái De Bù Quán Shì Guó Měi Pái Bīng Xiāng ? Nǚ Yán Jiū Shēng Gào Diàn Shāng Píng Tái ， Quán Guó Shǒu Lì " Sōu Suǒ Fú Wù " Wéi Yuē Àn Pàn Liǎo* (为啥搜出来的不全是国美牌冰箱？女研究生告电商平台，全国首例"搜索服务"违约案判了) [*Why the search is not all Gome brand refrigerators? A female graduate student sued the e-Commerce platform; the first case in China involving breach of contract for "search service" has been decided*], QIÁN JIĀNG WĂN BÀO (钱江晚报) [QIANJIANG EVENING NEWS] (2021), https://3g.163.com/dy/article/G2D205CM0512GTK3.html [https://perma.cc/EH2J-8VED] (describing a consumer dispute over Taobao search results in which the platform allegedly refused to explain the algorithm on trade secret grounds); Shàng Yóu Xīn Wén (上游新闻) [Upstream News], *Nǚ Hái Yīn " Hé Nán Rén " Yīng Pìn Bèi Jù Àn Kāi Tíng Zhè Jiāng Xǐ Lái Dēng Bèi Pàn Dào Qiàn Péi Cháng 1 Wàn Yuán* (女孩因"河南人"应聘被拒案开庭 浙江喜来登被判道歉赔偿1万元) [*The Trial Has Begun For The Case of a Girl Who Was Rejected for a Job Because She Was from Henan Province. The Sheraton Hotel in Zhejiang Was Ordered to Apologize and Pay 10,000 Yuan in Compensation*], PÉNG PÀI XĪN WÉN (澎湃新闻) [The Paper] (2019), https://www.thepaper.cn/newsDetail_forward_5064609 [https://perma.cc/T5DM-NLVN] (reporting a geographic discrimination case involving algorithmic resume screening); Yanjun Chao (曹彦君), *Dà Shù Jù Shā Shú, Wéi Shén Me Yuè Lái Yuè Sì Wú Jì Dàn?* (大数据杀熟，为什么越来越肆无忌惮?) [*Big Data Killing, Why Is It Getting More and More Unscrupulous?*], 21 SHÌ JÌ SHĀNG YÈ PÍNG LÙN (世纪商业评论) [21ST CENTURY BUS. REV.] (Dec. 21, 2020), https://m.21jingji.com/article/20201221/herald/c49f8dc4b126badd76b8c400b478e5c5_zaker.html [https://perma.cc/EC7H-9TXX] (discussing data-driven price discrimination in online hotel booking).

49. *See* Equal Credit Opportunity Act, 15 U.S.C. § 1691; Fair Credit Reporting Act, 15 U.S.C. §§ 1681e(b), 1681g, 1681i, 1681m; *see also* Hurley & Adebayo, *supra* note 16, at 170–75 (describing steps for using machine learning to build a big data model for monitoring credit scoring).

50. *See* Fair Credit Reporting Act, 15 U.S.C. §§ 1681c, 1681e(b), 1681g, 1681i; *see* Hurley & Adebayo, *supra* note 16, at 172–75 (addressing challenges of determining "creditworthiness"); *see* Persis Yu et al., *Big Data, a Big Disappointment for Scoring Consumer Credit Risk*, NAT'L CONSUMER L. CTR. at 5, 10 (Mar. 6, 2014), https://www.nclc.org/wp-content/uploads/2022/09/report-big-data.pdf [https://perma.cc/RQA2-VGL7] (describing roadblocks in accessing creditworthiness through various models).

give specific reasons for adverse action, either in the notice or upon timely request. Regulation B rejects explanations based solely on internal lending standards or failure to attain a qualifying credit score. However, reason-giving offers limited protection when consumers cannot identify the nontraditional data and correlations underlying their exclusion from credit.[51]

This article employs a comparative lens between China and the EU, not to rank regimes but to shed light on shared legal problems.[52] Chinese debates over algorithmic explanation have drawn on General Data Protection Regulation (GDPR) centered arguments, including the view that meaningful information about automated decision-making can support an *ex post*, decision specific right to explanation.[53] However, China's eventual approach does not simply reproduce that model.[54] Chinese scholarship has emphasized that the right to explanation operates differently in public law and private law settings:[55] individuals may claim a stronger *ex post* explanation from public authorities, while enterprise facing duties more often concern *ex ante* disclosure of basic algorithmic information and notice of potential adverse effects.[56] Recent comparative work on the Personal Information Protection Law (PIPL) similarly shows that the influence of the GDPR is evident but selective and institutionally mediated.[57] The PIPL borrows from GDPR style language while leaving room for distinct interpretations, enforcement, and regulatory priorities shaped by China's domestic

51. *See* Equal Credit Opportunity Act, 15 U.S.C. § 1691; 12 C.F.R. § 1002.9(b)(2) (requiring a statement of specific reasons for adverse action); *Consumer Financial Protection Circular 2022-03: Adverse Action Notification Requirements in Connection with Credit Decisions Based on Complex Algorithms*, CONSUMER FIN. PROT. BUREAU 2–4 (May 26, 2022), https://files.consumerfinance.gov/f/documents/cfpb_2022-03_circular_2022-05.pdf [https://perma.cc/S397-U2KH] [hereinafter CFPB 2022]; *Consumer Financial Protection Circular 2023-03: Adverse Action Notification Requirements and the Proper Use of the CFPB's Sample Forms Provided in Regulation B*, CONSUMER FIN. PROT. BUREAU 2–5 (Sep. 19, 2023), https://wayback.archive-it.org/23481/20250331183051/https://files.consumerfinance.gov/f/documents/cfpb_adverse_action_notice_circular_2023-09.pdf [https://perma.cc/7B9K-WRSQ] [hereinafter CFPB 2023], both Circulars were withdrawn on May 12, 2025 and are cited here as historical interpretations, *see* Interpretive Rules, Policy Statements, and Advisory Opinions; Withdrawal, 90 Fed. Reg. 20084 (May 12, 2025); *see* also Hurley & Adebayo, *supra* note 16, at 180–87.

52. *See infra* Part II (comparing automated credit scoring systems in China and the EU).

53. *See* Huanmin Lin & Hong Wu, *A Right to an Explanation of Algorithmic Decision-Making in China*, 52 HONG KONG L. J. 1163, 1166–67 (2022) [hereinafter Huanmin] (explaining an individual's "right to request an explanation from public authorities of specific algorithmic decisions *ex post*, as a logical corollary to their right to know under Chinese administrative law.").

54. *See id.* at 1189 (explaining the difficulties of algorithmic governance in considering overlapping values).

55. *See id.* at 1178–83 ("Distinguishing between Public and Private Domains to Configure Different Subtypes of the Right to Explanation.").

56. *See id.* at 1163 (arguing that China's approach to the right to explanation distinguishes between individual-government and individual-enterprise scenarios, and that a uniformly applicable right to explanation is difficult to design); *see also* Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"*, 38 AI MAG. 52, 52–57 (2017) (arguing that the GDPR supports a right to explanation for automated decision-making); Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIV. L. 233, 233–42 (2017) (arguing that GDPR Articles 13–15 should be read functionally to provide meaningful information about automated decisions).

57. *See* Wenlong Li & Jiahong Chen, *From Brussels Effect to Gravity Assists: Understanding the Evolution of the GDPR-Inspired Personal Information Protection Law in China*, 54 COMPUT. L. & SEC. REV. 1, 14–19 (2024) (explaining the GDPR-PIPL relationship within legal development in China).

legal and political setting.[58] This article, therefore, uses comparative legal analysis to examine how different regimes translate automated credit scoring into legal duties of explanation.[59] Its main comparison remains between the EU and China, since the two systems reveal both convergence and institutional divergence in the regulation of automated decision-making.[60]

The article also draws on the U.S. ECOA and FCRA as reference points for adverse action notices, credit report accuracy, dispute rights, and reasons for credit denials.[61] At the same time, the U.S. model remains incomplete.[62] Credit scoring formulas have historically been protected as trade secrets, and factor-based disclosures may tell consumers that a factor mattered without explaining how much it mattered, how the score was calculated, or how a consumer can meaningfully change the result.[63] The black box problem examined here is therefore not limited to deliberate concealment by platforms.[64] In credit scoring, opacity may arise from technical complexity, organizational fragmentation, trade secret protection, or the combined effect of these conditions.[65]

This article argues that at their core, rather than being mere arguments about access to information, disputes over automated credit scoring are consumer protection disputes regarding the allocation of financial opportunities and exclusion in AI governed credit markets.[66] The right to explanation acts as a torch in a black box market, making suspected bias, error, proxy discrimination, and unlawful exclusion visible enough to be contested and reviewed.[67] Part II examines credit scoring as an automation mechanism and a market governance

58. *See id.* at 19 (arguing that the PIPL reflects a nuanced interaction between GDPR influence and China's domestic political, cultural, and legal priorities rather than a straightforward Brussels Effect); Xiaodong Ding et al., *Same Text, Different Meaning: China's Risk-Based Approach to Data Protection*, 12 HUMAN. & SOC. SCI. COMMC'NS 1, 2 (2025) (arguing that China's risk-based approach gives formally similar data-protection language a distinct regulatory meaning).

59. *See infra* Part VI.B.

60. *See infra* Part II.A.

61. *See* Equal Credit Opportunity Act, 15 U.S.C. § 1691; 12 C.F.R. § 1002.9(b)(2) (requiring creditors to provide a statement of specific reasons for adverse action); Fair Credit Reporting Act, 15 U.S.C. §§ 1681e(b), 1681g, 1681i, 1681m (requiring reasonable procedures for accuracy, consumer access to file information, dispute rights, and adverse-action notices based on consumer-report information); *see* CFPB 2022, *supra* note 51, at 1 (clarifying that creditors using complex algorithms must still provide specific and accurate reasons for adverse action); CFPB 2023, *supra* note 51, at 1 (explaining that creditors may not rely on overly broad, generic, or inapplicable reasons when providing adverse-action notices).

62. *See supra* note 50 and accompanying text.

63. *See* Citron & Pasquale, *supra* note 24, at 22–25 (discussing the development of credit scoring, the limited explanatory value of key-factor disclosures, and the opacity created by trade secret protection over scoring formulas); Martha Poon, *Scorecards as Devices for Consumer Credit: The Case of Fair, Isaac and Company Incorporated*, 55 SOC. REV. 284, 297–300 (2007) [hereinafter Scorecards as Devices for Consumer Credit] (discussing the development of FICO's credit-scoring system as a market device for consumer credit); Martha Poon, *From New Deal Institutions to Capital Markets: Commercial Consumer Risk Scores and the Making of Subprime Mortgage Finance*, 34 ACCT. ORGS. & SOC'Y 654, 662 (2009) (analyzing credit scores as calculative risk-management technologies in mortgage finance); *see also* Fair Isaac Corp. v. Experian Info. Solutions, Inc., 645 F. Supp. 2d (D. Minn. 2009), aff'd, 650 F.3d 1139 (8th Cir. 2011) (describing disputes involving competing credit-scoring systems and confidential scoring-related materials).

64. *See* Citron & Pasquale, *supra* note 24, at 10 (describing the black box problem of credit scores).

65. *See id.* at 10–11 (discussing the complexity of opacity and credit scoring).

66. *See infra* Part V.C. (addressing the U.S. model for adverse action of credit explanation); *see generally infra* Part II (providing background on automated credit scoring).

67. *See infra* Part VI.B. for more information on variances on the right to explanation.

device.[68] Two key questions drive the examination, the first of which concerns classification: where there is nominal human involvement or if a human reviewer simply rubber stamps a system output, should the decision still be treated as automated for legal purposes?[69] The second question concerns materiality: what counts as a significant effect on individual rights and interests, and under what conditions should automated credit scoring be recognized as producing these effects?[70] This article's core claim is that in both cases, these effects must be treated as functional standards that are grounded in how decisions are made and how their impacts are experienced.[71]

In Part III, the analysis moves to the legal regulation of scoring, where the following question is asked: why do established remedies so often fail to give credit-market consumers a usable pathway to contest biased allocation in AI governed credit markets?[72] In contrast, automated scoring is usually embedded in standard-form terms and framed as a pre-contractual risk assessment or pricing infrastructure.[73] It enables lenders and platforms to present adverse outcomes as commercial discretion rather than legally reviewable reasoning while restricting consumers from any realistic prospect of informed consent or meaningful negotiation.[74] Meanwhile, in tort, the difficulty is rarely the absence of harm:[75] the structure of liability, including causation, fault, and unlawfulness, presupposes access to the model, data pipeline, and decision records held by the controller;[76] therefore, opacity converts evidential asymmetry into protection from liability.[77] Thus, the analysis proposes that a more explicit right to algorithmic explanation serves as the connecting principle linking legal effects to enforceable duties and effective contestability.[78]

Building on Part III, Part IV develops the consumer protection basis for a right to explanation in AI-governed credit markets by separating technical interpretability from legally usable justification.[79] It evaluates how explainable artificial intelligence (XAI) tools can support internal diagnostics, audit, and supervision, while showing why their limits, trade secret constraints, and model-level outputs do not by themselves give consumers reviewable reasons for adverse credit outcomes.[80] The article reframes responsibility from narrow causal attribution, which is put under strain by distributed agency and statistical inference,

68. *See infra* Part II for details on automated credit scoring.
69. *Infra* notes 152–55 and accompanying text.
70. *See infra* notes 187–89 and accompanying text (discussing automated credit scoring affects).
71. *See infra* Part II for a detailed response to these two key questions.
72. *See infra* Part III for the legal effects of regulating credit scoring.
73. *See infra* Part III.A; *infra* notes 222–25 and accompanying text (describing the contractual framework).
74. *See infra* Part III.A; *infra* notes 261-62 for the contractual framework of credit scoring disputes.
75. *See infra* Part III.B; *infra* note 263 for the evidentiary challenges that arise from tort remedies.
76. *See infra* Part III.B; *infra* note 264 and accompanying text (noting what claimant is required to prove).
77. *See infra* note 271 and accompanying text (noting the difficulty of proving causation for a claimant).
78. *Infra* notes 321–22 and accompanying text.
79. *See infra* Section IV (addressing the explainability, responsibility, and trade secrecy of AI governed credit markets and their algorithms).
80. *Infra* Part IV.A, *infra* notes 336, 340 and accompanying text.

to power and answerability.[81] Therefore, the right to explanation serves as a form of procedural fairness and contestability rather than a thin information right.[82] Finally, the article addresses trade secret objections, which are often used to refuse disclosure in practice,[83] by advocating for a layered model of explanation and third-party review such that consumers, regulators, and courts receive the level of detail needed without forcing the public release of proprietary models.[84]

Next, Part V examines whether existing law clearly confers a right to algorithmic explanation.[85] It moves from access, notice, and adverse action reasons to the stronger question of whether legal regimes make explanation practically contestable.[86] Some scholars advocate synthesizing disparate provisions into a de facto right, particularly through purposive readings of Article 22 and Recital 71 of the GDPR.[87] Other scholars contend that the law, at most, imposes a duty to inform rather than a substantive right to individualized justification.[88] The United States ECOA and FCRA provide a useful contrast, since adverse-action notices, credit-file access, and dispute rights show how reason-giving can become actionable while also exposing the limits of factor-based disclosure.[89] This article argues that the doctrinal disagreement is most effectively resolved by recognizing that the legal system already contains the ingredients of explanation, but that a more explicit and detailed legal construction is necessary to make explanation operational and contestable in credit-scoring disputes.[90] A further—cautiously optimistic—development is that the EU Artificial Intelligence Act holds the potential to shake up this impasse.[91] Crucially, by naming the deployer as the duty bearer, it gives consumers a statutory basis for who to sue and what to demand,[92] which may spur more cases and better protect consumers'

81. *Infra* Part IV.B; *infra* notes 382–87 and accompanying text.

82. *Infra* Part IV.B*; infra* notes 390, 394–96 and accompanying text.

83. *See infra* Part IV.C, notes 405–07 (discussing right to explanation potential conflicts with trade secrets).

84. *Infra* notes 414–17 and accompanying text.

85. *See infra* Part V (discussing whether algorithmic explanation rights exist in the law).

86. *See infra* Part V.A–C (addressing practicality of contestability standard).

87. *See* Bryan Casey et al., *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L. J. 145, 180–84 (2019) (discussing how enhancing the interpretability of algorithms through mechanisms like the "right to explanation" and data protection impact assessments (DPIAs) is essential for fostering accountability, transparency); *see also* Selbst & Powles, *supra* note 56, at 242 (discussing how the GDPR's provisions inherently support a "right to explanation" for automated decisions, emphasizing the importance of meaningful information about the logic involved to ensure transparency, accountability, and fairness in algorithmic processes).

88. Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 76, 77 (2017) (discussing how the GDPR does not explicitly mandate a "right to explanation" for automated decision-making, but rather establishes a "right to be informed" about system functionality and logic, highlighting the challenges and limitations in achieving full transparency and accountability in algorithmic processes).

89. *See supra* note 50–51 and accompanying text (discussing the ECOA and FCRA).

90. *Infra* Part VI and accompanying text.

91. *See infra* Part VI.A.1; *infra* note 679 (explaining the entitlement to an explanation under the EU AI Act).

92. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act), art. 86. 2024 O.J. (L 1689) [hereinafter EU AI Act].

epistemic justice in practice.[93] Simultaneously, by introducing this right without an accompanying avenue of redress for affected individuals, the AI Act may foster the belief that no additional legal solutions are required.[94] Against this backdrop, this article's comparative analysis suggests that the EU's AI Act, the EU's GDPR, and China's PIPL do not yet supply a complete and workable pathway from opacity to contestability in credit-scoring disputes.[95] The proposal, therefore, specifies the duty bearer;[96] the procedural pathway for requesting, receiving, and challenging explanations;[97] and a graduated framework for explanation content.[98] This article adopts a sober stance toward techno-optimism and legal formalism.[99] Automation alone is no guarantee of fairness while disclosure alone is no guarantee of accountability.[100] What matters is how legal standards define automation,[101] how legal effects are recognized across doctrinal fields,[102] and how the explanation is structured to enable contestation.[103]

## II. AUTOMATED CREDIT SCORING

Automated credit scoring should first be situated within the economic logic of credit markets.[104] Traditional underwriting relies heavily on repayment history, income, assets, and other structured financial records.[105] Such records can exclude borrowers who are creditworthy but have thin files, are underbanked, are informally employed, or are otherwise invisible to conventional scorecards.[106] Machine-learning credit models address that problem by processing larger, more varied datasets, detecting nonlinear relationships, and extracting predictive signals from alternative data, weak financial signals, and digital footprints.[107] Alternative-data scoring may therefore produce consumer-welfare benefits by lowering underwriting costs, expanding lending capacity, improving

93. *See* Jackie Kay et al., *Epistemic Injustice in Generative AI*, PROC. SEVENTH AAAI/ACM CONF. ON AI, ETHICS, & SOC. 684, 691–92 (2024) (arguing that transparency can remedy the epistemic injustice caused by opacity in AI systems).

94. Simona Demkova, *The AI Act's Right to Explanation: A Plea for an Integrated Remedy*, MEDIALAWS (Oct. 31, 2024), https://www.medialaws.eu/the-ai-acts-right-to-explanation-a-plea-for-an-integrated-remedy/ [https://perma.cc/2XUB-799G].

95. *Id.*

96. *Infra* Part VI.

97. *Id.*

98. *Id.*

99. *Id.*

100. Joshua A. Kroll*, Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems*, PROC. 2021 ACM CONF. FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 758 (2021).

101. *Id.*

102. *Id.*

103. *Id.*

104. Satyajit Chatterjee et al., *A Quantitative Theory of the Credit Score*, 91 ECONOMETRICA 1803 (2023).

105. *See* Hurley & Adebayo, *supra* note 16, at 162–63 ("The FICO score, for instance, principally looks at a consumer's payment history, the amounts she owes, the length of her credit history, new credit, and types of credit she uses.").

106. *Id*. at 154-55.

107. *See* Khandani et al., *supra* note 9, at 1–3, 12–19 (finding that integrated transaction, credit-bureau, and account balance data reveal nonlinear predictive patterns beyond traditional credit scores).

default prediction, and identifying candidates who might otherwise be misclassified as high risk or unscorable.[108]

However, better prediction can alter the distribution of winners and losers because more granular risk classification may generate greater dispersion in prices, approvals, and loan terms.[109] Fintech lending may reduce some forms of human discretion, yet discrimination can persist through pricing, proxy variables, and model-mediated underwriting.[110] Large language models add an emerging layer to this infrastructure by extracting soft information from borrower narratives, loan applications, customer communications, bank statements, business descriptions, compliance files, and other unstructured materials.[111]

Thus, automated credit scoring should also be understood not simply as a consumer-facing prediction tool, but as part of a bank's regulated credit-decision infrastructure.[112] In the EU, the European Banking Authority treats automated models used in creditworthiness assessment and credit decision-making as matters of loan-origination governance.[113] Institutions are required to understand model inputs, assumptions, limitations, outputs, documentation, monitoring, and the role of automated outputs in credit decisions.[114] The prudential framing becomes clearer under the EU Capital Requirements Regulation.[115] Use of the

108. *See* Berg et al., *supra* note 9; Chunxiao Li et al., *The Effect of AI-Enabled Credit Scoring on Financial Inclusion: Evidence from an Underserved Population*, 48 MIS Q. 1803, 1803–34 (2024) (examining how AI-enabled credit scoring affects approval, default, and utilization among underserved borrowers); World Bank & Int'l Comm. on Credit Reporting [ICCR], *The Use of Alternative Data in Credit Risk Assessment*, at 14–20, 42–46 (2024), https://documents1.worldbank.org/curated/en/099031325132018527/pdf/P179614-3e01b947-cbae-41e4-85dd-2905b6187932.pdf [https://perma.cc/U6J8-Y4FP] [hereinafter World Bank] (discussing how alternative data may support credit access for consumers and MSMEs lacking conventional credit histories).

109. *See* Fuster et al., *supra* note 12, at 13–17 (finding that machine learning can improve prediction while unevenly distributing benefits across borrower groups).

110. *See* Bartlett et al., *supra* note 15 (finding that discrimination persists in fintech-era lending); *see* Hurley & Adebayo, *supra* note 16, at 156–61, 178–83 (warning that big-data credit scoring may rely on nontraditional signals that consumers may struggle to understand or challenge).

111. *See* Li et al., *supra* note 10 (explaining that LLMs can transform unstructured financial text into decision-relevant information); *see* also Yi et al., *supra* note 11 (surveying AI applications in accounting and finance); Mario Sanz-Guerrero & Javier Arroyo, *Credit Risk Meets Large Language Models: Building a Risk Indicator from Loan Descriptions in P2P Lending*, 28 INTELIGENCIA ARTIFICIAL 220, 221 (2025); Yufei Xia et al., *Credit Scoring Model for Fintech Lending: An Integration of Large Language Models and FocalPoly Loss*, 41 INT'L J. FORECASTING 894, 894–96 (2025); U.S. DEP'T OF THE TREASURY, ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICE: USES, OPPORTUNITIES, AND RISKS OF ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES 16 (2024), https://home.treasury.gov/system/files/136/AI-in-Financial-Services-Report.pdf [https://perma.cc/ASR2-ZEJ7] [hereinafter U.S. Dep't of the Treasury] (identifying document extraction, document querying, risk identification, and compliance support as financial-sector AI use cases).

112. *See* 12 C.F.R. § 1002.2(p) (defining a credit scoring system); *see generally* BD. OF GOVERNORS OF THE FED. RSRV. SYS., SR 26-2, SUPERVISORY GUIDANCE ON MODEL RISK MANAGEMENT (2026), https://www.federalreserve.gov/supervisionreg/srletters/SR2602a1.pdf [https://perma.cc/8R23-Z7FL] [hereinafter Guidance on MRM] (subjecting such model development, validation, monitoring, and governance to supervisory obligations).

113. EURO. BANKING AUTH., EBA/GL/2020/06, *Guidelines on Loan Origination and Monitoring*, ¶ 41, https://www.eba.europa.eu/sites/default/files/document_library/Publications/Guidelines/2020/Guidelines%20on%20loan%20origination%20and%20monitoring/884283/EBA%20GL%202020%2006%20Final%20Report%20on%20GL%20on%20loan%20origination%20and%20monitoring.pdf [https://perma.cc/3UPK-JX8J].

114. *Id.* ¶¶ 33, 41, 53–55, 121, 141.

115. *See* Regulation (EU) No. 575/2013 of the European Parliament and of the Council, arts. 143–44, 2013 O.J. (L 176) (requiring institutions to obtain supervisory permission before using internal models to calculate credit risk exposure).

internal ratings-based approach requires supervisory permission, and rating systems must be embedded in credit approval, risk management, internal capital allocation, and corporate governance, instead of being treated as freestanding statistical instruments.[116]

Comparable supervisory logic appears in the United States and China.[117] U.S. banking regulation governs internal ratings-based and advanced approaches for risk-weighted assets, while federal model-risk guidance treats model development, use, validation, monitoring, governance, controls, and proportionality to the bank's model-risk profile as supervisory concerns.[118] Chinese banking regulation is aligned with the same prudential structure.[119] The Commercial Bank Capital Management Measures require commercial banks using the internal ratings-based approach to satisfy regulatory conditions and pass supervisory acceptance,[120] and the 2024 rules on advanced capital-measurement methods define advanced methods as internal models for measuring risk and regulatory capital, including the credit-risk internal ratings-based approach, with model outputs to be applied in business decision-making and risk management.[121]

In addition, risk appetite is not external to the model.[122] It is transmitted through board-approved limits, capital targets, approval thresholds, pricing bands, portfolio concentrations, probability-of-default cutoffs, and delegated authority.[123] JPMorgan Chase's Risk Committee, for example, annually reviews and approves the firm's quantitative and qualitative risk appetite and receives reports when results exceed or are forecasted to exceed that appetite.[124] Citi's Risk Management Committee Charter requires the committee to review Citi's risk profile against the board-approved risk appetite and oversee key policies

116. *Id.*

117. *Infra* notes 118–21 and accompanying text.

118. 12 C.F.R. pt. 217, subpt. E.; Guidance on MRM, *supra* note 112.

119. Guó Jiā Jīn Róng Jiān Dū Guǎn Lǐ Zǒng Jú Guī Zhāng (国家金融监督管理总局规章) [Nat'l Fin. Regul. Admin.], *Shāng Yè Yín Xíng Zī Běn Guǎn Lǐ Bàn Fǎ* (商业银行资本管理办法) [*Measures for the Capital Management of Commercial Banks*] (promulgated Nov. 1, 2023, effective Jan. 1, 2024), https://www.nfra.gov.cn/cn/view/pages/rulesDetail.html?docId=1134197 [https://perma.cc/8Q28-HNUY] [hereinafter Measures for the Capital Management of Commercial Banks].

120. *Id.*

121. *Id.* art. 49; Guó Jiā Jīn Róng Jiān Dū Guǎn Lǐ Zǒng Jú (国家金融监督管理总局) [National Financial Regulatory Administration], *Guó Jiā Jīn Róng Jiān Dū Guǎn Lǐ Zǒng Jú Guān Yú Yìn Fā Shāng Yè Yín Xíng Shí Shī Zī Běn Jì Liáng Gāo Jí Fāng Fǎ Shěn Qǐng Jí Yàn Shōu Guī Dìng De Tōng Zhī* (国家金融监督管理总局关于印发商业银行实施资本计量高级方法申请及验收规定的通知) [*Provisions on the Application for and Acceptance of Advanced Approaches for Capital Measurement by Commercial Banks*], arts. 2, 15 (Nov. 1, 2024), https://www.nfra.gov.cn/cn/view/pages/governmentDetail.html?docId=1184708&generaltype=1&itemId=861 [https://perma.cc/HZ7A-NSN4].

122. Measures for the Capital Management of Commercial Banks, *supra* note 119, art. 130.

123. *Id.* arts. 140, 143.

124. *See Risk Committee*, JPMORGAN CHASE & CO., https://www.jpmorganchase.com/about/governance/committees/risk-committee [https://perma.cc/BX34-KP8V] (last visited Oct. 29, 2026) ("Review and approve the Firm's quantitative and qualitative risk appetite annually. Receive reports on results against defined risk appetite. The Risk Committee Chair shall be notified promptly if Firm Risk Appetite results have exceeded or are forecasted to exceed risk appetite.").

that establish risk limits.[125] Its shareholder materials then translate that governance structure into capital-allocation language, explaining that management was rigorous about risk and returns and made "deliberate decisions about where to deploy capital and resources and where to step back."[126]

Automated credit scoring does not merely predict default risk;[127] it operationalizes a bank's permitted level of credit expansion, expected return, capital consumption, and tolerance for loss within a supervisory framework that links models, governance, capital, and managerial accountability.[128] Therefore, credit scoring is a quantified description of an individual's credit position.[129] It uses data-driven techniques to rank and differentiate between subjects using predefined or empirically derived criteria to provide an ostensibly objective basis for downstream decision-making.[130] Automated credit scoring can be unpacked along two axes, namely, automation and credit scoring.[131] Their interaction reveals a distinct operating model for credit evaluation organizations.[132] Before turning to any particular regulatory scheme, it is therefore necessary to clarify how existing laws define the operational features of automation itself, particularly how the GDPR and PIPL[133] construct automation as a legally relevant mode of processing and decision-making.[134]

### *A. Automated decision-making*

Algorithmic innovation can improve how credit markets process information and allocate resources by enabling lenders to process larger datasets, detect nonlinear relationships, and identify credit signals that traditional underwriting may miss.[135] However, the resulting opacity is not reducible to deliberate

125. *Risk Management Committee Charter*, CITIGROUP INC., at 3 (Feb. 11, 2026), https://www.citigroup.com/rcs/citigpa/storage/public/citigroup-rmc-charter.pdf [https://perma.cc/M5VJ-CSAA].

126. *Id.*; Jane Fraser, *Letter to Shareholders*, Citigroup Inc. (2026), https://www.citigroup.com/global/investors/annual-reports-and-proxy-statements/2026/annual-report/letter-to-shareholders [https://perma.cc/U35Z-J2KG].

127. *See* BD. OF GOVERNORS OF THE FED. RSRV. SYS., REPORT TO THE CONGRESS ON CREDIT SCORING AND ITS EFFECTS ON THE AVAILABILITY AND AFFORDABILITY OF CREDIT 5, (Aug. 2007), https://www.federalreserve.gov/boarddocs/rptcongress/creditscore/creditscore.pdf [https://perma.cc/RD4J-NQUK] [hereinafter Report to Congress on Credit Scoring] (explaining that credit scoring facilitates risk-based pricing, broadens creditors' access to capital markets and reduces funding costs).

128. *Supra* notes 118–123 and accompanying text.

129. *See* Report to Congress on Credit Scoring, *supra* note 127, at 1 ("Credit scoring is a statistical technology that quantifies the credit risk posed by a prospective or current borrower.").

130. *See* Citron & Pasquale, *supra* note 24, at 31 (explaining credit scoring purports "to scientifically rank individuals' creditworthiness with extraordinary precision").

131. Nikita Aggarwal, *Machine Learning, Big Data and the Regulation of Consumer Credit Markets: The Case of Algorithmic Credit Scoring*, *in* AUTONOMOUS SYS. & THE L. (2019).

132. *Id.*

133. GEREN XINXI BAOHU FA (个人信息保护法) [PERSONAL INFORMATION PROTECTION LAW] (promulgated by Standing Comm. Nat'l People's Cong., Aug. 20, 2021, effective Nov. 1, 2021) [hereinafter PIPL]; Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119), 1 [hereinafter GDPR].

134. *Infra* notes 143–44 and accompanying text.

135. *See* Khandani et al., *supra* note 9; Julapa Jagtiani & Catharine Lemieux, *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform*, 48 FIN. MGMT.

concealment by platforms or lenders.[136] At the data level, alternative data, data-quality problems, and cross-contextual inferences make it difficult to trace the evidentiary basis of a score.[137] At the organizational level, scoring chains may divide control among lenders, platforms, model vendors, and cloud providers, leaving no single actor visibly responsible for the whole decision pipeline.[138] At the strategic level, firms may invoke proprietary model logic, trade secrecy, or contractual discretion to resist disclosure of information needed for challenge.[139] The legal problem is not simply that consumers know less, since consumers cannot determine whether an adverse credit outcome reflects an accurate risk assessment, a data error, unlawful discrimination, or discretionary exclusion.[140] Rather, opacity reallocates power by widening asymmetries in information, cognition, and the consumers' practical ability to claim or contest entitlements.[141] Individual credit status is therefore exposed to a combined dominance of informational control and automated decision practice.[142]

In Chinese law, automation is a regulatory trigger.[143] Article 73(2) of the PIPL offers the following definition: "'Automated decision-making' refers to the activities of automatically analyzing and evaluating personal behaviors, hobbies, or economic, health, and credit status, among others, through computer programs, and making decisions."[144] In credit markets, this definition fits the dominant data–model–and scenario architecture of personal credit evaluation, where the model refers to an automated computation that produces, recalibrates, and refreshes the credit market consumer's score.[145] The scoring system is

1009, 1010–13 (2019) (examining how alternative data and machine learning affect fintech lending and credit access).

136. Citron & Pasquale, *supra* note 24.

137. Hurley & Adebayo, *supra* note 16; *see* World Bank, *supra* note 108 (discussing the use of alternative data in credit assessment and its implications for access, accuracy, and discrimination).

138. *See* U.S. DEP'T OF THE TREASURY, *supra* note 111, at 25–31 (discussing third-party dependence, model governance, and risk-management issues in financial-sector AI); *see also* World Bank *supra* note 108, at 26–28 (discussing collaborations between banks, fintech agencies, and credit reporting agencies); *see generally* Yujin Zhou et al., *How Well Can Large Language Model Agents Simulate Complex Legal Dispute Resolution?*, A.I. & L., May 2026 (showing, in a multi-agent legal simulation setting, that LLM agents' outputs are shaped by role assignment, interaction design, model-level configuration, and procedural setting, thereby illustrating how responsibility can become distributed across system architecture rather than residing in a single decision-maker).

139. *See* Citron & Pasquale , *supra* note 24, at 10–11, 17 (discussing trade-secret protection and the opacity of credit-scoring formulas).

140. *See* Hurley & Adebayo, *supra* note 16, at 178–80 (discussing issues of accuracy with large data sets used by credit scoring systems); *see* Citron & Pasquale, *supra* note 24, at 11–17 (discussing how consumers lack information on credit scoring due to opacity).

141. *See* Huq, *infra* note 352, at 682–83 (arguing that machine-learning tools deepen power asymmetries between individuals and organizations, and that rights alone do not fix that imbalance).

142. *See* Citron & Pasquale, *supra* note 24 at 17, 21 (discussing how consumers have limited information about credit scores and the potential risks automation poses).

143. *See* PIPL, art. 24(3) (establishing legal consequences for automated decision-making).

144. PIPL, art. 73(2).

145. *See* Jiang Minghui (姜明辉) et al., *Geren Xinyong Pingfen Moxing de Fazhan ji Youhua Suanfa Fenxi* (个人信用评分模型的发展及优化算法分析) [*Research on Algorithms Development and Optimization for Personal Credit Scoring*], 5 HAERBIN GONGYE DAXUE XUEBAO (哈尔滨工业大学学报) [J. HARBIN INST. TECH.] 40 (2015) (discussing how the evolution of personal credit scoring models from statistical methods to AI techniques necessitates a focus on explainability to ensure model transparency, fairness, and regulatory compliance, addressing challenges like sample validity, indicator rationality, and model applicability).

typically controlled by credit evaluation organizations or platform-enabled third parties;[146] meanwhile, the consumer only experiences the output.[147] Article 22 of the GDPR exposes the core interpretive fault lines: whether there is a decision and whether it is based solely on automated processing.[148] A sound textual reading of Article 22(1) begins from its object: *"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."*[149] The intricacy of this provision is salient because liability can be influenced by small changes in how its connective terms are construed.[150] A counterargument is that controllers may seek to sidestep the solely automated threshold by inserting a nominal human reviewer.[151] However, Article 29 Data Protection Working Party (WP29)[152] treats designs where the human merely rubber-stamps an automated output as insufficient and lacking genuine authority or competence to change the result.[153] While this position is administratively convenient, it can be analytically misleading because it assumes a simple and linear pipeline.[154] As human intervention and a decision's significance can be stratified across stages and outcomes, an earlier automated decision may remain determinative of the outcome even if a human appears at the end.[155]

Additionally, in *SCHUFA*, the referring court treats a probability value, or a statistically generated score expressing the likelihood of future credit performance, as practically decisive because it "normally determines whether and how" a lender contracts.[156] It reasons that, if Article 22(1) only applies at the

146. *See infra* note 209 and accompanying text (Civil Code art. 1030, referring to "credit information handlers such as credit rating agencies"); *see infra* note 229 and accompanying text (illustrating platform-enabled third-party control of scoring through the Zhima Credit/Taobao arrangement).

147. *See* Citron & Pasquale, *supra* note 24, at 17 (discussing that credit scores must be disclosed, but the reasoning behind the score remains undisclosed).

148. *See infra* notes 149–184 and accompanying text (discussing Article 22(1)'s requirement of a "decision" "based solely on automated processing").

149. GDPR, art. 22(1).

150. *See* Luca Tosoni, *The Right to Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation*, 11 INT'L DATA PRIV. L. 145, 146–47 (2021) [hereinafter Luca] (explaining that whether Article 22(1) is a prohibition or a right to object turns on small differences in how its wording is read).

151. *See* Binns & Veale, *infra* note 155, at 320–21, 330 (explaining that a token human reviewer may not satisfy the "solely automated" standard, and could exploit this by inserting a nominal reviewer to avoid Article 22).

152. *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, ART. 29 DATA PROT. WORKING PARTY, at 5, 21 (2018) [hereinafter WP29] ("To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data").

153. *Id.*

154. *See id.* at 28 (stating that human controllers should audit data sets frequently and continuously).

155. *See* Reuben Binns & Michael Veale, *Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR*, 11 INT'L DATA PRIV. L. 319, 320–21 (2021) (discussing how the interpretation of Article 22 of the GDPR, which restricts solely automated decisions with legal or significant effects, faces ambiguities in multi-stage profiling systems, highlighting the need for clearer guidelines on human intervention and the significance of decision outcomes; "As human intervention and a decision's significance can be stratified across stages and outcomes.").

156. Case C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, ¶ 22.

lender's final act, a "lacuna in legal protection" would arise because the scoring entity would not count as adopting automated decision-making for Article 15(1)(h), whereas the lender receiving the score would not have the additional information needed to explain it.[157] The Court therefore treats the concept of "decision" in Article 22(1) as broad enough to capture a creditworthiness probability value.[158] It also confirms that the automated establishment of such a value constitutes automated individual decision-making where a third party relies on it to establish, implement, or terminate the contractual relationship.[159] This effect-based reading allows Article 22 to function as an anti-circumvention norm because the trigger conditions remain textually disciplined by the requirement of legal effects or similarly significant effects.[160]

The same interpretive posture is seen in *Dun & Bradstreet Austria*.[161] The Court notes that the versions of Article 15(1)(h) pull in different directions;[162] while some stress intelligibility, others value, functionality, relevance, or importance; therefore, it treats those meanings as complementary.[163] This makes it harder for controllers to rely on formal labels or semantic minimalism, as the right is read by reference to its function.[164] The Court correspondingly reads the "logic involved" as covering procedure and principles and insists on disclosure that supports contestation, including "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[165] Notably, the provision targets a decision and not the upstream processing, and the phrase "*based on*" signals a dependency relation rather than an identity claim.[166] Therefore, the law does not require that "the processing that leads to" an outcome be treated as the outcome itself for the purposes of legal qualification.[167] In this regard, decision-making is most accurately understood as a distinct normative act of taking a stance toward a person, which can be separated from preparatory or supporting stages that structure, inform, or operationalize the eventual stance, even when those stages are automated.[168] A bank's refusal to execute a transaction can thus be treated as the

157. *Id.* ¶ 23.

158. *See id.* ¶ 46 (discussing that 'decision' is so broad that it includes a creditworthiness probability value).

159. *See id.* ¶ 73 (discussing probability value qualifying as automated decision-making).

160. *See id.* ¶¶ 43–46, 61–63 (discussing the broad coverage of 'decision' and the risks of interpreting Article 22 narrowly).

161. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117.

162. *See id.* ¶ 40 (discussing the variety of interpretations of Article 15(1)(h)).

163. *See id.* ¶¶ 40–43 (discussing that versions of Article 15(1)(h) are complementary).

164. *See id.* ¶¶ 46, 66 (discussing that Article 15(1)(h) implications on the controller).

165. *Id.* ¶ 6 ("The existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.").

166. *See* Bygrave, *Automated Profiling*, *infra* note 168, at 19 (explaining that decisions must be distinguished from "other processes … that can prepare the way for, or head off, formal decision making").

167. Luca, *supra* note 150, at 153.

168. *See* Lee A. Bygrave, *Automated profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, 17 COMP. L. SEC. REP. 17, 19 (2001) [hereinafter "Bygrave, *Automated Profiling"*] (discussing how Article 15 of the EC Data Protection Directive addresses automated profiling by granting individuals the right not to be subject to solely automated decisions significantly affecting them, emphasizing the need for transparency and explainability in algorithmic decision-making to protect personal

legally salient decision, whereas an antecedent anti-money laundering monitoring pipeline remains an automated process that conditions the refusal.[169] Thus, responsibility is attached to the measure that produces legal effects without requiring every antecedent to be reclassified as a decision.[170]

These cases can be understood as converging on a broad, effects-sensitive understanding of decision that reduces the scope for formalistic evasion in multi-actor systems.[171] A counterargument is that an expansive approach risks rendering the word "solely" in Article 22(1) superfluous.[172] A settled interpretive principle—recognized in the EU and elsewhere[173]—holds that courts should, so far as possible, read legislation so that no word or phrase is left without operative effect, consistent with the presumption of the legislator's rationality.[174] On that view, it is difficult to see why Article 22(1) would single out decisions "based solely on automated processing" if the provision is ultimately meant to reach cases involving meaningful human involvement.[175] In this regard, if Article 22(1) can be triggered even where a human is present somewhere in the chain, the text would cease to distinguish automated from hybrid decision governance and create undue uncertainty.[176] The cleaner view is that the presence of a human at the end of the workflow is not the issue but whether the relevant decision point is, in substance, based solely on automated processing.[177] When the final outcome reflects a meaningful human re-evaluation, it may fall outside the scope of Article 22(1).[178] However, earlier stages can constitute solely automated decisions if actors in the later stage treat their outputs as being effectively

integrity and dignity; "Difficulties may arise sometimes in distinguishing decisions from other processes . . . that can prepare the way for, or head off, formal decision making"); *see also* Lee A. Bygrave, *Article 22 Automated Individual Decision-Making, Including Profiling*, *in* THE EU GENERAL DATA PROTECTION REGULATION (GDPR) 522, 532 (Christopher Kuner ed., Lee A. Bygrave ed., & Christopher Docksey ed., 2020) [hereinafter "Bygrave, *Article 22"*] (defining a decision as "a particular attitude or stance…taken towards a person" with "a degree of binding effect…likely to be acted upon"); Case C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, ¶ 37.

169. *See* Bygrave, *Automated Profiling*, *supra* note 168, at 19–21 (discussing automation and bank loans).

170. *See supra* notes 156, 169 and accompanying text (establishing the legal-effects trigger and the decision/preparatory process distinction, respectively).

171. *See supra* notes 156–70 and accompanying text (discussing *SHUFA*'s holding that "decision" under Article 22(1) reaches a credit-scoring probability value, and *Dun and Bradstreet Austria*'s holding that "meaningful information" requires disclosure of procedure and principles supporting contestation).

172. *See* Binns & Veale, *supra* note 155, at 320 (discussing 'solely' interpretation).

173. Koen Lenaerts & José Gutiérrez-Fons, *To Say What the Law of the EU Is: Methods of Interpretation and the European Court of Justice*, 20 COLUM. J. EUR. L. 3, 13–14 (2014); *see also* Duncan v. Walker, 533 U.S. 167, 174 (2001) (quoting Market Co. v. Hoffman, 101 U.S. 112, 115 (1879)) ("[A] statute ought, upon the whole, to be so construed that, if it can be prevented, no clause, sentence, or word shall be superfluous, void, or insignificant.").

174. *See* Lenaerts & Gutiérrez-Fons, *supra* note 173, at 14 ("As a token of rationality, the EU legislator must also avoid useless duplication. Accordingly, no provision of EU law should be redundant. Instead, each and every provision of that law must be interpreted in light of its 'effet utile.'").

175. *See* Bygrave, *Article 22*, *supra* note 168, at 532 (discussing 'solely' in Article 22(1)).

176. *See id.* at 532–33 (discussing human interaction with decision-making process).

177. *See supra* notes 151–55 and accompanying text (discussing human interaction in the automated decision-making process).

178. *See* WP29, *supra* note 152 ("To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data.").

determinative and do not reopen the assessment.[179] This reading preserves the internal logic of Article 22(1) while accounting for multi-stage systems.[180] A further objection is the administrative burden, as courts may need to examine internal workflows, contractual practices, and evidentiary patterns of reliance.[181] However, as per *Dun & Bradstreet Austria*, complexity cannot lower the transparency threshold and right of access is necessary to enable the exercise of downstream rights.[182] This implies that administrability concerns cannot justify an interpretation that makes rights contingent on how a controller allocates tasks in a socio-technical chain.[183] Therefore, the preferable balance is to treat broad decisions as an anti-circumvention move that targets practical effects while retaining the limiting conditions, including the requirement of legal or similarly significant effects.[184]

### *B. Hybrid legal interest*

This article defines automated credit scoring as a practice wherein credit management bodies and digital platforms, after obtaining the subject's consent, use their personal credit-related information such as identity data, transaction records, and social behavioral signals to generate a score through a preset algorithmic model.[185] The normative stakes lie in how a numerical proxy becomes the basis for allocating opportunities and burdens.[186] The nature of credit is commonly framed along two dimensions: an evaluative personality interest related to the social appraisal of economic reliability and an intangible property interest reflecting economic value through its impact on transactional opportunities.[187] As automated credit scores are embedded in platform-service filtering and differential pricing practices, they directly influence equality of opportunity in access to services and market participation,[188] whereas data error and model bias can diffuse harms across linked decisions through a domino effect that is difficult to isolate and prove.[189]

179. *See supra* note 155, 160 and accompanying text (discussing the importance of earlier, automated stages in regards to assessments).

180. *See* WP29, *supra* note 152, at 21("The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing.").

181. *See* Asymina Aza, *Scores as Decisions? Article 22 GDPR and the Judgment of the CJEU in SCHUFA Holding (Scoring) in the Labour Conte*xt, 53 INDUS. L.J. 840, 856 (2024) (noting that determining the degree to which human decision-makers are able to depart from automated scores or assessments "may not be as unambiguous" as it was in *SCHUFA* itself).

182. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 54, 61.

183. *See supra* notes 181–82 and accompanying text (establishing that administrative burden and complexity should not justify minimizing transparency).

184. *See* Luca, *supra* note 150 (discussing generally the best interpretation of Article 22).

185. *See* Huanmin, *supra* note 53, at 1168–70 (discussing the issues of automated algorithms)..

186. *Id.* at 1180.

187. *See* Wu Handong (吴汉东), *Lùn Xìn Yòng Quán* (论信用权) [*On the Right to Credit*], 1 FAXUE (法学) [LAWSCI.] 41, 42 (2001) (discussing how the legal protection of credit rights as an intangible property right necessitates clear definitions and direct safeguards).

188. Li et al., *supra* note 108, at 1823–24.

189. *See* Huanmin, *supra* note 53, at 1184–85.

This duality has surfaced in the debate over recognizing an independent credit right, as the legal form determines when liability is triggered and what remedies are available.[190] One line of scholarship calls for the right to an autonomous commercial personality, treating the characterization of infringement and the rules governing damages as the backbone of protection.[191] Other approaches have sought to fold credit into a general rights, duties, and liability framework that connects civil law to credit governance or called for a dedicated commercial credit regime to govern information collection and evaluation, along with the design of incentive and sanction mechanisms.[192] A further strand rejects the need for a standalone credit right, construing the core harm in scoring disputes as being reputational and relational rather than proprietary and enabling an expanded reading of reputation protection to capture the wrong without proliferating rights categories.[193] This route is attractive due to its procedural economy, as courts may focus on how an adverse score communicates negative appraisal and drives exclusion rather than aiming to resolve the contested boundaries of a distinct credit entitlement.[194] However, this strategy risks structural distortion: it systematically sidelines the property-like dimension of credit and treats the loss of future access and transaction opportunities as consequential rather than a cognizable interest warranting tailored corrective relief.[195] It also fits poorly with automated scoring.[196] Even where reputational harm is acknowledged, asymmetry impedes identification of the operative rule, proxy, or data defect that produced the score.[197] Therefore, the remedies tend to stop at *ex post* moral vindication rather than ensuring *ex ante* contestability and model-fitting correction.[198]

190. For an overview of arguments for and against the establishment of credit as an individual right, *see* Marek Hudon, *Should Access to Credit be a Right?*, 84 J. BUS. ETHICS 17, 18–25 (2009).

191. *See* Wang Liming (王利明), *Rén Gé Quán Fǎ Yán Jiū* (人格权法研究) [*Research on Personality Rights Law*], 中国人民大学出版社 [CHINA RENMIN UNIV. PRESS], 268–72 (2018) (arguing that personality rights possess an autonomous commercial dimension, the protection of which should be grounded in infringement analysis and damages rules).

192. .*See* Zhao Xudong (赵旭东), *Shāng Shì Xìn Yòng De Jiè Dìng Yǔ Zhì Dù Gòu Chéng* (商事信用的界定与制度构成) [*Definition and Institutional Composition of Commercial Credit*], 6 浙江工商大学学报 [J. ZHEJIANG GONGSHANG UNIV.] 5, 8–9 (2019) (discussing how the GDPR does not explicitly mandate a "right to explanation" for automated decision-making, but rather establishes a "right to be informed" about system functionality and logic, highlighting the challenges and limitations in achieving full transparency and accountability in algorithmic processes).

193. *See* Zhang Jihong (张继红), *Gè Rén Xìn Yòng Quán Yì Bǎo Hù De Sī Fǎ Kùn Jìng Jí Qí Jiě Jué Zhī Dào ——— Yǐ Gè Rén Xìn Yòng Quán Yì Jiū Fēn De Sī Fǎ Àn Lì* (个人信用权益保护的司法困境及其解决之道———以个人信用权益纠纷的司法案例 (2009－2017) 为研究对象) [*Judicial Dilemma of the Protection of Personal Credit Rights and Its Solutions—Taking the Judicial Case of Disputes Over Personal Credit Rights (2009–2017) as the Object of Study*], 33 FǍ XUÉ LÙN TÁN (法学论坛) [L. F.] 138 (2018) [hereinafter Judicial Dilemma] (discussing how the legal ambiguity surrounding personal credit rights in China leads to inconsistent judicial rulings).

194. *Id*.

195. *See* Joao Marinotti, *Possessing Intangibles*, 116 NW. UNIV. L. REV. 1227, 1258–60 (2022) (criticizing the treatment of interference with intangible interests as merely consequential harm and emphasizing the property-like character of such interests).

196. *See* Judicial Dilemma, *supra* note 193 (discussing the issues with this strand of thinking).

197. Nilantha Prasad et.al, *A Survey of Cyber Threat Attribution: Challenges, Techniques, and Future Directions*, COMPUTERS & SEC. (2025).

198. *Id.*

Moreover, the GDPR's algorithmic information duties in Articles 13–15,[199] grounded in the lawfulness, fairness, and transparency principle in Article 5(1),[200] operate as behavioral norms across the decision-making lifecycle rather than as purely *ex post* repair devices.[201] This behavioral and process-facing logic sits in tension with the personality rights architecture in China's Civil Code,[202] which primarily protects recognized absolute rights through claim rights aimed at two core functions: filling already materialized personality harm and restoring the completeness of the infringed right.[203] In addition, its preventive effects remain tethered to an *ex post* remedial grammar and the independently recognized personality rights confirmed by Article 110.[204] Because personal information interests are an emergent legal interest arising from contemporary information practices,[205] their right density and infringement modalities do not map neatly onto the absolute right paradigm;[206] therefore, courts face practical and conceptual difficulties in defining a restored complete state.[207] The Civil Code situates credit evaluation relationships within the protection of personal information and does not confer explicit independent status on a credit right, commonly understood as an interests-protection model as opposed to a rights-creation model.[208] Article 1030 provides that "[t]he provisions of this Book on the protection of personal information and the relevant provisions of other laws and administrative regulations shall be applied to the relationship between persons of the civil law and the credit information processors such as a credit reporting agency."[209] The Civil Code's personality and request rights and liabilities are oriented

199. *See, e.g.*, Bygrave, *Automated Profiling*, *supra* note 168 (discussing Article 15 responsibilities); *see also* Commission Regulation 2016/679, 2016 O.J. (L 119) 1, art. 13–15 (EU) [hereinafter GDPR].

200. *See* GDPR, art. 5.

201. *See* Casey et al., *supra* note 87, at 176–80 (arguing that the information duties in Articles 13–15 GDPR, grounded in the lawfulness, fairness, and transparency principle, operate ex ante across the decision-making lifecycle rather than merely as ex post remedies).

202. Zhonghua Renmin Gongheguo Minfadian (中华人民共和国民法典) [Civil Code of the People's Republic of China] (promulgated by the Nat'l People's Cong., May 28, 2020, effective Jan. 1, 2021), art. 989–1001 (China) [hereinafter "Chinese Civil Code"].

203. *See* MO ZHANG, CONTEMPORARY CHINESE LAW AND LEGAL SYSTEM (Cambridge Univ. Press 2026) (describing the theoretical and statutory split between Book IV (Personality Rights) and Book VII (Tort Liability) of the Civil Code).

204. Chinese Civil Code, art. 110.

205. WANG LIMING & XIONG BINGWAN, THE MAKING OF THE CHINESE CIVIL CODE: PROMISES AND PERSISTENT PROBLEMS, at 41–42 (Hao Jiang and Pietro Sirena eds. 2023) [hereinafter Making of the Chinese Civil Code].

206. *See, e.g.*, CHRIS D. LINEBAUGH & STEVE P. MULLIGAN, CONG. RSCH. SERV., IF11207, DATA PROTECTION AND PRIVACY LAW: AN INTRODUCTION (2022) (discussing the complex patchwork of federal and state laws).

207. *See, e.g.*, *id.* (mentioning how the Supreme Court has interpreted privacy).

208. *See* Cheng Xiao (程啸), *Lùn Wǒ Guó Mín Fǎ Diǎn Zhōng Gè Rén Xìn Xī Quán Yì De Xìng Zh* (论我国民法典中个人信息权益的性质) [*On the Nature of the Right and Interest of Personal Information in the Civil Code of China*], 8 Zhèng Zhì Yǔ Fǎ (政治与法律) [POL. SCI. & L.] 2, 7 (2020) [hereinafter Cheng] (discussing the legal recognition of personal information as a personality right in China's Civil Code); *see also* Zhou Hanhua (周汉华), Gè Rén Xìn Xī Bǎo Hù De Fǎ Lǜ Dìng Wèi (个人信息保护的法律定位) [*The Legal Orientation of Personal Information Protection*], 37 FǍ SHĀNG YÁN JIŪ (法商研究) [STUD. L. & BUS.] 65, 71 (2020) (discussing the legal ambiguity in China regarding the nature of personal information protection—whether as a right or interest).

209. Chinese Civil Code, art. 1030.

toward filling realized harm and restoring the completeness of protected absolute rights;[210] meanwhile, their remedial set is expressed through Articles 179 and 995, which do not map cleanly onto the relational, processing-centered nature of personal information interests.[211] Likewise, deriving an algorithmic explanation injunction from the special personality injunction procedure under Article 997 is textually and structurally difficult when the interest at stake is personal information rather than a recognized personality right.[212] By the same logic, it is difficult, based on the text and the underlying remedial design, to extend the special personality injunction mechanism in Article 997 to an algorithmic explanation injunction where the protected interest is personal information.[213] This difficulty is reinforced in credit-scoring disputes because fault is often harder to construct in a principled manner, given that the enquiry does not stop at whether the actor breached a general duty of care recognized by ordinary social standards but must also address whether the actor complied with the sector-specific reasonable review obligations imposed by Article 23(1) of the *Regulation on the Administration of Credit Investigation Industry* and Article 9 of the *Measures for the Administration of the Credit Reference Business*.[214]

## III. REGULATING CREDIT SCORING AND LEGAL EFFECTS

Automated credit scoring disputes are, in practice, disputes about how data-driven inference is allowed to allocate opportunities and liabilities.[215] Before developing a remedy-facing entitlement that targets opacity at its source, it is therefore necessary to clarify how these existing doctrinal routes operate and where they predictably fail.[216] On the one hand, the contract frame tends to reduce contestability to standard-term authorization and platform autonomy.[217] The case turns on whether the platform had contractual power to score, adjust, and recalibrate.[218] On the other hand, tort-based routes often stall on proof.[219] Even where reputation and related doctrines offer entry points, claimants still face an

210. *See id.* art. 995 (giving a person whose personality rights are infringed the right to request the actor "stop the infringement, remove the nuisance, eliminate the danger, eliminate the adverse effects, rehabilitate his reputation, or extend apologies").

211. *Id.* art. 179, 995.

212. *See id.* art. 997 (contrasting personal information structure versus art. 995's personality right).

213. *See id*. (explaining the special personality injunction mechanism).

214. *See* Cheng Xiao (程啸), *supra* note 208 ("Where an information provider provides credit information for a credit reference agency, the credit reference agency shall develop relevant rules, to conduct necessary review of the information provider's information source, information quality, and information security, and authorization of the information subject.").

215. *See, e.g.*, Murali Malempati, *Ethical considerations and regulatory challenges in data-driven finance and credit assessment*, DEEPSCIENCE (2025) (describing the ethical issues of using data-driven finance).

216. *Id.*

217. Tariq K. Alhasan et. al, *Governing Automated Credit After Explainability: From Transparency to Contestability*, SOC, SCI. & HUMAN. OPEN (2026).

218. *Id.*

219. *See, e.g.*, 15 U.S.C. § 1681h(e) (heavily preempting state-law tort claims).

evidentiary wall.[220] They must prove illegality, causation, and fault without access to model logic, data features, or validation records.[221]

### *A. The contractual framework*

The legal governance of automated credit scoring turns, at the outset, on classification.[222] A contractual approach treats platform rules as functionally analogous to standard-form contractual terms on the basis that platform service protocols operate as standing offers to an indefinite class and that user assent is manifested through click-through or continued use, thereby forming a click contract.[223] Accordingly, the platform's capacity to structure transaction formation, pricing, performance, and liability supports it being treated as a transactional party and, depending on the content, distinct platform rules being located within corresponding contractual relations.[224] The same logic has been used to conceptualize credit-scoring rules as allocating a risk-bearing function akin to a guarantee structure, where the platform leverages the score to underwrite trust and performance between counterparties.[225]

This channeling effect is visible in how courts routinely classify automated credit-scoring disputes as contract cases and then decide them through the preselected cause-of-action templates of service, sale, or lease, which privileges contractual authorization over substantive contestability.[226] When platforms score users, they typically rely on the parties' contractual relationship, most often a platform service agreement;[227] thus, disputes are readily framed as contractual disputes and adjudicated within fixed doctrinal categories.[228] In *Huang Gang v. Zhejiang Taobao Network Co.*[229], the claimant argued that damage to an electronic product during transit was not his fault and that a 99-point

220. *See, e.g.*, *id.* (finding under United States law, consumers must prove the entity acted with malice or willful intent to injure the consumer).

221. *Id.*

222. *See, e.g.*, Philipp Hacker & Maximilian Eber, *The Future of Credit Underwriting and Insurance Under the EU AI Act: Implications for Europe and Beyond*, HARV. SCI. DATA R., 2025 (discussing one method of classification).

223. *See* Xia Qingfeng (夏庆锋), *Wǎng Luò Hé Tóng Zhōng Bù Zhèng Dāng Gé Shì Hé Tóng De Jiū Zhèng Guī Zé* (网络合同中不正当格式合同的纠正规则) [*Correction Rules for Unfair Standard Terms in Online Contracts*], 2 JIĀNG HUÁI LÙN TÁN (江淮论坛) [JIANGHUAI TRIB.] 130, 134 (2020) (discussing the prevalence of unfair terms in online contracts, due to inadequate consumer consent mechanisms).

224. *See* Yang Lixin (杨立新), *Wǎng Luò Jiāo Yì Fǎ Lǜ Guān Xì Gòu Zào* (网络交易法律关系构造) [*The Construction of Legal Relations in Online Trading*], 2 ZHŌNG GUÓ SHÈ HUÌ KĒ XUÉ (中国社会科学) [SOC. SCI. CHINA] 114, 117–25 (2016) (Discussing the complex legal relationships in online transactions, involving multiple parties).

225. *See* Jiang Daxing (蒋大兴) & Wang Shoujie (王首杰), *Gòng Xiǎng Jīng Jì De Fǎ Lǜ Guī Zhì* (共享经济的法律规制) [*Legal Regulation of the Sharing Economy*], 9 ZHŌNG GUÓ SHÈ HUÌ KĒ XUÉ (中国社会科学) [SOC. SCI. CHINA] 155 (2017) (discussing the innovative nature of the sharing economy necessitates regulatory adaptations).

226. *See, e.g.*, 黄刚诉浙江淘宝网络有限公司 [Huang Gang v. Zhejiang Taobao Network Co., Ltd.], (2019) Zhe 01 Min Zhong No. 7214, at 3–4, Hangzhou Interm. People's Ct. (China) (holding that the platform's adjustment of Zhima Credit scores constituted a lawful and reasonable performance of the Taobao platform service agreement).

227. *Id.*

228. *Id.*

229. *Id.*

deduction in *Zhima Credit (芝麻信用)* was unreasonable, and sought restoration.[230] The Court held that a network service contract existed under the Taobao platform agreement and treated the scorer's adjustment as a lawful and reasonable performance of the contract.[231] In *Wang Zhizong v. Zhejiang Taobao Network Co.*[232], the Court followed a similar route in treating a downgrade from five stars to two as within the platform's autonomous management and rejecting damages because the act was not found to violate mandatory law or consumer rights.[233] The pattern matters because most judgments directly apply platform rules or user agreement clauses, which makes the litigation hinge on whether the platform had contractual power to score and adjust, as opposed to it hinging on whether the score was produced through reliable data processing and a contestable reasoning process.[234] This is not a minor omission.[235] It effectively converts a dispute about the legality and fairness of automated inference into a dispute about the scope of a standard term, which predictably favors the party that both drafts the term and controls the information required to rebut it.[236]

The contractual framework becomes even thinner when the dispute concerns dynamic scoring systems.[237] This is because the apparent technical neutrality of score recalibration can mask a systematic monopoly over rule interpretation and thus defeat the very possibility of informed consent.[238] In the *Nanhan*

230. *Id.*

231. *Id.*

232. *See* 王至宗诉浙江淘宝网络有限公司 [Wang Zhizong v. Zhejiang Taobao Network Co., Ltd.], (2018) Jing 0114 Min Chu No. 14773, at 6–7, Changping Dist. People's Ct. of Beijing (China) (holding that the Taobao platform service agreement constituted a network service contract and that the platform's adjustment of the user's credit rating fell within lawful and reasonable contractual performance).

233. *Id.*

234. *Cf.* Ye Ming (叶明) & Qiu Weiqi (邱威棋), *supra* note 27, at 127 (discussing how contract-law remedies for platform credit-scoring disputes are undermined by privity limitations–users can typically sue only the nominal scoring entity, not the platform–and by the difficulty of quantifying score-related losses; not addressing the litigation-framing point asserted in text).

235. *Cf.* Citron & Pasquale, *supra* note 24, at 19 (arguing that the absence of legal accountability for algorithmic decision-makers risks creating "a new feudal order of unaccountable reputational intermediaries"); *see also* Ye Ming (叶明) & Qiu Weiqi (邱威棋), *supra* note 27, at 120–22 (2023) (arguing that nondisclosure of scoring standards creates credit anxiety, lack of supervision can facilitate manipulation, and inability to verify scoring results can undermine the digital trust mechanism).

236. *See* Xia Qingfeng (夏庆锋), *supra* note 223, at 134 (noting that the absence of a developed legal framework for online contracts, combined with judicial tolerance for reduced bargaining, gives platforms–as drafters of standard-form terms–a structural advantage over users, who often cannot ascertain how far click-through or browse-wrap terms constrain their rights).

237. *See* Liu Quan (刘权), Ping Tai Gui Ze de Fa Zhi Bian Jie (平台规则的法治边界) [*The Rule-of-Law Boundaries of Platform Rules*], 4 ZHONGGUO SHEHUI KEXUE (中国社会科学) [Soc. Sci. China] (2026) (arguing that platform rules generated and revised after the initial contract is formed do not fit the "pre-drafted term" paradigm that standard-form-contract doctrine presupposes, since platforms continuously add and revise rules—including scoring rules—after users have already joined the platform).

238. *See* Huang Shaokun (黄绍坤), *Ping Tai Gui Ze de He Fa Xing Kong Zhi* (平台规则的合法性控制) [*Legitimacy Control of Platform Rules*], BEIJING LIGONG DAXUE XUEBAO (SHEHUI KEXUE BAN) (北京理工大学学报（社会科学版) [J. BEIJING INST. TECH (SOC. SCI. ED.)] (2023) (arguing that platform operators' technical and economic dominance gives them a form of private power over users, under which formally-required notice and informed consent mechanisms increasingly become mere formalities incapable of performing their intended function).

*Korean BBQ* case[239], the handover gap between old and new credit-scoring rules directly resulted in an erroneous store closure.[240] Although Article 34 of the *E-Commerce Law of the PRC* requires platforms to publicize rule changes seven days in advance,[241] the platform treated a backend push of a New Credit Score Interpretation as sufficient notice, while the merchant in practice faced a cognitive lag in understanding and responding to the change.[242] The Court found a breach but did not negate the validity of the rule change, exposing the limitations of formal compliance reviews.[243] While a system push may satisfy a paper notice logic,[244] it would fail the functional question of whether notice was effective in high-throughput settings where merchants handle hundreds of daily orders and cannot realistically translate opaque scoring guidance into protective action.[245] Besides the imperfect nature of notice, formal notice can operate as a legal shield for unilateral recalibration;[246] thus, the platform's interpretive monopoly remains intact and harms caused by transition design are treated as incidental mishaps rather than governance failures demanding structural remedies.[247]

A deeper form of contractual limitation appears where institutions invoke contract terms and business necessity to resist any duty to justify scoring logic, because that stance allows automated inference to control legal effects while

239. *See* Jiamusi City Xiangyang (West) District Nanhan Korean BBQ Restaurant v. Shanghai Rajax Information Technology Co., Ltd., (2018) Hu 02 Min Zhong No. 955, Shanghai Second Intermediate People's Ct. (China) (the handover gap between old and new credit-scoring rules directly resulted in an erroneous store closure).

240. *Id.*

241. Zhongua Renmin Gongheguo Dianzi Shangwu Fa (中华人民共和国电子商务法) [E-Commerce Law of the People's Republic of China] (promulgated by the Standing Comm. Nat'l People's Cong., Aug. 31, 2018, effective Jan. 1, 2019), art. 34 (China) [hereinafter E-Commerce Law].

242. *See Nanhan*, Hu 02 Min Zhong No. 955 (noting that the platform relied on a backend push of the revised credit-scoring interpretation as its notice to the merchant).

243. *Id.* (finding that the platform's notice practice breached its disclosure obligations but declining to invalidate the underlying credit-scoring rule change).

244. *See* Xia Qingfend (夏庆锋), *Wangluo Hetong Geshi Tiaokuan Shuoming Yiwu de Lvxing Xiaci yu Wanshan Cuoshi* (网络合同格式条款提示说明义务的履行瑕疵与完善措施) [*Deficiencies in the Performance of Notice and Explanation Obligations for Standard Terms in Online Contracts, and Improvement Measures*], 16 QINGHUA FAXUE (清华法学) [Tsinghua Univ. L.J.] 118 (2022) [hereinafter Xia, *Notice Deficiencies*]; *see also* E-Commerce Law, art. 34.

245. *Guifan Pingtai Shoufei, Shanghu Ruzhu Deng! Waimai "Xinggui" Zhengqiu Yijian* (规范平台收费、商户入驻等！外卖"新规"征求意见) [*Standardizing Platform Fees, Merchant Onboarding, and More: Public Comment Sought on New Food-Delivery Rules*], ZHŌNG XĪN WĂNG (中新网) [China News Serv. ] (Sep. 24, 2025), https://www.chinanews.com.cn/cj/2025/09-24/10488084.shtml [https://perma.cc/BF2V-S79J] (quoting Zhang Xin, Professor, Univ. of Int'l Bus. & Econ. Sch. of L., on the "order-overload" phenomenon in which merchant service capacity is instantaneously overwhelmed during order peaks).

246. *See* Huang Shaokun (黄绍坤), *supra* note 238 (arguing that formal compliance mechanisms increasingly function as mere formalities detached from substantive accountability); *see also Nanhan*, Hu 02 Min Zhong No. 955 (illustrating this dynamic, where the court found a notice-related breach but declined to invalidate the underlying rule change); *see also* E-Commerce Law, art. 34.

247. *See* Huang Shaokun (黄绍坤), *supra* note 238 (on platforms' private-power monopoly over rule interpretation); *see also* Liu Quan (刘权), *supra* note 237 (arguing that treating platform rules as ordinary contract terms limits correction to case-by-case contract remedies, foreclosing the structural governance tools—transparency mandates, filing/review mechanisms, participatory rule-making—that a genuine governance failure would require); *see also* E-Commerce Law, art. 34.

remaining evidentially insulated.[248] In *Liu Shuxing*,[249] when the claimant's credit card was disabled, the bank relied on contractual clauses to refuse explanation of the scoring logic, and the Court's acceptance of business needs effectively exempted the institution from a justificatory obligation that would otherwise enable challenge and correction.[250] The reasoning problem is sharper than a simple deference to business judgment.[251] Once the model is treated as a protected internal risk tool, the subject cannot prove error, discrimination, or misapplication;[252] therefore, rights in form exist without a pathway to vindication in practice.[253] The same structural asymmetry distorts how courts assess whether unilateral change is reasonable.[254] In the *Nanhan Korean BBQ*[255] case, the platform failed to clarify the territorial scope of the new scoring regime yet demanded compliance from merchants who had not meaningfully experienced the new system,[256] a posture that sits uneasily with the consent requirement for contract modification under Article 543 of the Civil Code, where "a contract may be modified if the parties reach a consensus through consultation."[257] In *Liu Shuxing*[258], the Court treated high aggregate credit as being equivalent to deteriorated creditworthiness and discounted the probative value of historical performance, which reads less as neutral evaluation than a judicial strategy to stabilize the enforceability of standard terms in the face of an explanatory deficit.[259]

248. *See* Citron & Pasquale, *supra* note 24, at 5 (noting that scorers "refuse to reveal the method and logic of their predictive systems," treating the scoring process itself as immune from external justification); *see also* Liu, *infra* note 249.

249. *See* Liu Shuxing v. Beijing Bank Co., Ltd. Credit Card Ctr. (刘树星诉北京银行股份有限公司信用卡中心), (2022) Jing 0113 Min Chu No. 2067, Shunyi Dist. People's Ct. of Beijing (China) (accepting the bank's reliance on standard-form credit card contract clauses and business necessity to justify disabling the claimant's credit card without disclosing scoring logic, and treating aggregate credit exposure as evidence of deteriorated creditworthiness notwithstanding the claimant's long-term repayment history).

250. *Id.*

251. *Cf.* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 633 (2017) (arguing that traditional legal tools for reviewing decisions—designed to oversee human decision-makers—fail when the decision-maker is an algorithm, because affected individuals and oversight bodies alike lack any way to verify that the actual decision process matches the announced rules).

252. *See id.* (arguing that when an automated decision system is shielded from disclosure, affected individuals lack any means to verify whether the actual decision process contained errors, discriminatory correlations, or a misapplication of the announced rule).

253. *Id.*; *cf.* Roscoe Pound, *Law in Books and Law in Action*, 44 AM. L. REV., 12 (1910) (on the broader jurisprudential distinction between formally declared rights and their practical vindication).

254. Chinese Civil Code, art. 543; *see also* Lin Huanmin (林洹民), *Shuzi Fuwu Hetong Danfang Biangeng Quan zhi Guizhi* (数字服务合同单方变更权之规制) [*Regulation of the Right of Unilateral Modification in Digital Service Contracts*], XIANDAI FAXUE (现代法学) [MODERN L. SCI.] (2023, No. 2) (arguing that a platform's unilateral-modification clause must be assessed for reasonableness under Civil Code Article 497(2), and that the platform itself bears the burden of justifying the clause's reasonableness—specifying the circumstances, content, and impact of the change and obtaining the counterparty's separate consent—rather than the burden falling on the affected party to disprove it).

255. *See* Jiamusi City Xiangyang (West) District Nanhan Korean BBQ Restaurant v. Shanghai Rajax Information Technology Co., Ltd., (2018) Hu 02 Min Zhong No. 955, Shanghai Second Intermediate People's Ct. (China), at 4.

256. *Id.*

257. Chinese Civil Code, art. 543.

258. Liu Shuxing v. Beijing Bank Co., Ltd. Credit Card Ctr. (刘树星诉北京银行股份有限公司信用卡中心), (2022) Jing 0113 Min Chu No. 2067, Shunyi Dist. People's Ct. of Beijing (China), at 4.

259. *Id.*

Consequently, the remedial horizon for score subjects is reduced to exit or acquiescence because the contract frame supplies a binary logic of either "accept" or "terminate."[260] However, the real question is how to contest an automated inference that produces legal effects in contract performance and, in practice, determines whether a person can continue to transact on equal terms.[261] This limitation demonstrates why the next section must develop a remedy-facing entitlement that targets opacity at its source, rather than leaving legality to turn on formal assent and *ex post* contractual rationalization.[262]

### *B. Evidentiary challenges in tort remedies*

Evidentiary difficulties are not accidental in tort-based challenges to automated credit scoring.[263] This is because the claimant is typically asked to prove illegality, harm, causation, and fault while the decisive information is located within an opaque, defendant-controlled scoring pipeline.[264] Within the internal logic of the Civil Code, the starting point appears supportive.[265] Article 1024 treats credit as a form of social evaluation, placing it within the protective confines of reputation,[266] whereas Article 1029 grants the right to raise an objection, correction, and deletion, signaling that distorted credit information can be a legally cognizable wrong.[267] However, in practice, the pathway from these entitlements to successful tort relief is fragile as credit scores often generate losses that are economic and opportunity-based rather than expressively reputational[268] and because the evidential burden in reputation torts is calibrated for public denigration and not for contested inferences produced by automated systems.[269] The

260. Friedrich Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629, 632 (1943).

261. *See* PIPL, art. 24 (2021) (China) ("[r]equiring transparency and fairness in automated decision-making and granting individuals the right to request an explanation and refuse decisions based solely on automated decision-making when the decision significantly affects their rights and interests").

262. *Id.*

263. Chinese Civil Code, Art. 1165 (2020); Kroll et al., *supra* note 251; Citron & Pasquale, *supra* note 24.

264. Chinese Civil Code, art. 1165 (fault-liability principle: "行为人因过错侵害他人民事权益造成损害的，应当承担侵权责任"—a person who causes harm to another's civil rights and interests through fault shall bear tort liability); *see* Chinese Civil Code, art. 1029 (providing that a person whose credit evaluation is improperly handled may request correction or deletion of inaccurate information); Kroll et al., *supra* note 251; Citron & Pasquale, *supra* note 24.

265. Chinese Civil Code, arts. 1024, 1029.

266. Chinese Civil Code, art. 1024

267. Chinese Civil Code, art. 1029.

268. *See* Citron & Pasquale, *supra* note 24; *see also* Leonardo Gambacorta et al., *How Do Machine Learning and Non-Traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm*, 73 J. FIN. STABILITY, at 78 (2024) (finding that machine-learning credit-scoring models built on non-traditional data more accurately predict borrower losses and defaults than traditional models, particularly during periods of credit-market stress, underscoring that such scores function to forecast financial risk and access to credit rather than to certify an individual's reputation).

269. *Cf. Mingyu Qinquan Ruhe Rending? Yanlun Ziyou yu Mingyu Qinquan de Jiexian Zai Nali? Yi Wen Du Dong!* (名誉侵权如何认定？言论自由与名誉侵权的界限在哪里？一文读懂！) [*How Is Reputation Infringement Determined? Where Is the Line Between Freedom of Expression and Reputation Infringement? A Full Explainer*], THE PAPER (澎湃新闻) (Aug. 19, 2021), https://m.thepaper.cn/newsDetail_forward_14123082 [https://perma.cc/58WM-P6XH] (last visited Sep. 14, 2026) (explaining that PRC reputation-tort liability

claimant must therefore litigate the score's normative meaning and the decision process' technical correctness but without having access to the model features, weights, or training logic that can substantiate an allegation of error or bias.[270] This is a structural asymmetry that turns fault and causation into near formalities rather than contestable questions.[271]

The gap between reputation doctrine and score-based harm is illustrated by cases where courts refuse to consider that internal platform-scoring measures lower the claimant's social evaluation, even when the claimant plausibly suffers real economic consequences from the score's signaling function.[272] In *Wang Pei*[273], an online game platform banned the claimant's account for alleged cheating and their forum credit rating fell to one star.[274] The claimant argued that their trading partners would infer that he was a low credit cheater, damaging his in-game interests.[275] Thus, he brought a defamation suit.[276] The Court treated the context as decisive, as in an environment dominated by strangers, a platform's internal sanction would be unlikely to map the account onto the claimant as an individual;[277] thus, the sanction was not characterized as a publicly derogatory evaluation akin to the typical reputation tort.[278] Therefore, relief (e.g., a public apology and moral damages) was denied.[279] The evidential lesson here is that even when the score operates as a market signal that affects transaction opportunities, courts may nevertheless require a recognizable reputational mechanism for public disparagement, narrowing the scope of legally cognizable harm and making causation speculative because besides establishing that the score

requires the defamatory content be disclosed to and known by a third party before a decrease in social evaluation can be established).

270. *See* Kroll et al., *supra* note 251; *see also* Citron & Pasquale, *supra* note 24 (on the lack of access to model features, weights, or training logic needed to substantiate a claim of algorithmic error or bias); *see also* Mo Chen et al., *On the Transparency of the Credit Reporting System in China*, 10 HUMANITIES & SOC. SCI. COMMC'NS 678 (2023) (discussing transparency limitations in digitally implemented Chinese credit-reporting systems).

271. *See* Jason Jia-Xi Wu, *Algorithmic Fairness in Consumer Credit Underwriting: Towards a Harm-Based Framework for AI Fair Lending*, 21 BERKELEY BUS. L. J. 65 (2024) (discussing how the framework for algorithmic fairness in consumer credit underwriting must incorporate transparency and explainability to mitigate potential harms, ensuring that AI-driven lending decisions are both fair and accountable to consumers).

272. Wang Pei v. Shenzhen Tencent Computer Sys. Co., Ltd. (王培诉深圳市腾讯计算机系统有限公司), (2018) Jing 01 Min Xia Zhong No. 374, Beijing First Interm. People's Ct. (China) [hereinafter Wang Pei]; Kroll et al., *supra* note 251.

273. *Wang Pei*, Jing 01 Min Xia Zhong No. 374 (involving an online game platform's ban of the claimant's account for alleged cheating, which also reduced the claimant's forum credit rating and allegedly impaired his in-game trading interests).

274. *Id.*

275. *Id.*

276. *See Canghai, DN* (玩家因违规被处罚，起诉腾讯被驳回) [*A DNF Player Was Punished for Violating Rules, and His Lawsuit Against Tencent Was Dismissed*], IT Home (Nov. 6, 2019), https://www.ithome.com/0/454/959.htm [https://perma.cc/JEY3-4XK5]

277. *Id.*

278. *See Wǎng Yóu Zhàng Hào Zāo Fēng Hào Néng Fǒu Qǐ Sù Qīn Fàn Míng Yù Quán?* (网游账号遭封号能否起诉侵犯名誉权？) [*Can You Sue for Defamation If Your Online Game Account Is Banned?*], JĪNG FĂ WĂNG SHÌ (京法网事) [BEIJING LEGAL AFFAIRS NETWORK] (July 15, 2019), https://www.bj148.org/sa1/yasf1/201907/t20190715_1520572.html [https://perma.cc/98RQ-B2YB] (reporting that Wang sued for defamation but his claim is distinct from a traditional defamation claim).

279. *Id.*

changed, the claimant must prove that, as a result, identifiable third parties formed a negative social evaluation of the claimant as a civil subject.[280] This doctrinal filter, while being coherent within traditional personality protection, systematically under-recognizes how the scoring system converts data processing into distributive outcomes by affecting eligibility and future opportunities that cannot easily be translated into conventional reputational injury.[281]

Even when courts are willing to engage more substantively, the fault-based structure of Article 1165 and the defendant's trade secrecy posture may prevent the dispute from reaching the algorithmic center such that the adjudication delivers case-specific correction without generating an audit-based pathway for scoring governance.[282] In *Gu Linxia v. Shanghai Hantao Information Consulting Co.*[283], the claimant challenged a platform's credit-score deduction and related sanctions as having harmed her reputation and honor.[284] She argued that her interactions were normal social communications and did not constitute the disruption of platform order alleged, relying on the user integrity code rules.[285] The Court subjected the claimant's conduct to substantive review by re-examining, through human assessment, whether her behavior fit the platform's rules, which can be considered an attempt to restore procedural fairness where automated classification is disputed.[286] However, this technique does not solve the evidentiary problem, which is salient for systemic governance.[287] The platform invoked a commercial secrecy clause in Section 3.4 of the User Integrity Code[288], claiming that the disruption model and its algorithm were core business secrets, and refused disclosure of the technical basis.[289] This stance meant that the claimant could not know the logic by which her conduct was characterized as a violation in the first place.[290] Based on these elements, the claimant had to effectively prove that the score deduction and sanction were unlawful or unjustified, that reputational harm had followed, and that the defendant was at fault;[291] however,

280. *Id.*

281. *See id.* (explaining the difficulty in establishing reputational harm).

282. Chinese Civil Code, art. 1165.

283. *See* Gu Linxia v. Shanghai Hantao Info. Consulting Co., Ltd. (顾琳霞诉上海汉涛信息咨询有限公司), Hu 0107 Min Chu No. 16697, Putuo Dist. People's Ct. of Shanghai (China) (2020) [hereinafter *Gu Linxia*] (involving a challenge to a platform's credit-score deduction and related sanctions, where the claimant argued that ordinary social interactions did not constitute the alleged disruption of platform order and that the measures harmed her reputation and honor).

284. *Id*. at 2.

285. *Id*. at 2.

286. *See id*. at 3.

287. *Cf. id*. at 3.

288. *Id.* at 3; *see also* Meituan (美团), *Meituan Yonghu Chengxin Gongyue* (美团用户诚信公约) [*Meituan User Integrity Convention*] § 3.4, (May 20, 2019), https://rules-center.meituan.com/rules-detail/7 [https://perma.cc/98N4-7JMQ] (describing that the defense invoked the "trade secret clause," asserting that the "platform order disruption model" system and its algorithms constitute core trade secrets. Refusal to disclose the specific algorithmic basis prevented the plaintiff from understanding the rationale behind the decision that their actions constituted a violation).

289. *See* Gu Linxia v. Shanghai Hantao Info. Consulting Co., Ltd. (顾琳霞诉上海汉涛信息咨询有限公司), Hu 0107 Min Chu No. 16697, Putuo Dist. People's Ct. of Shanghai (China) (2020).

290. *Id*.

291. *Id.*

opacity converted these requirements into an evidential *cul de sac*, given that the most probative evidence was withheld and the claimant could not generate substitutes with comparable diagnostic power.[292]

In addition, while products liability is sometimes treated as an alternative evidential shortcut,[293] it is doctrinally and practically ill-suited to algorithmic credit scoring.[294] The product regime is designed for tangible, replicable industrial goods, whereas scoring algorithms are software-based, dynamic, and continuously updated;[295] thus, whether an algorithmic scoring system is even a product for liability purposes is disputed.[296] Even if characterized as a product, the relevant issue is rarely a manufacturing flaw.[297] More plausibly, it is an implicit design defect expressed as inferential gaps, biased data, or inadequate warnings;[298] however, those defects are difficult to specify *ex ante* and even more difficult to connect to a specific negative score, because the causal pathway runs through training distributions, parameterization choices, and context-dependent deployment practices.[299] Furthermore, developers may invoke a development risk defense under Article 41 of the *Law of the PRC on Product Quality*, further increasing the claimant's burden by reframing structural uncertainty as lawful innovation risk.[300] The claimant would typically lack the technical capacity to prove defect and causation at the required level.[301] Finally, even when causation could be demonstrated, product liability tends to compensate realized losses,[302] whereas credit-scoring errors typically hinder expected interests by foreclosing future opportunities to obtain loans, insurance, or employment, which are central to governance concerns but are difficult to monetize and often discounted within traditional make-whole compensation logics.[303] This suggests that tort- and product-based remedies leave a predictable evidential gap in automated credit-scoring disputes.[304] The claimant is asked to prove what only the defendant can

292. *Id.*

293. *See* Yang Fan (杨帆), *Xinyong Pingfen Suanfa Zhili: Suanfa Guizhi yu Chanpin Zeren de Rongtong* (信用评分算法治理：算法规制与产品责任的融通) [*Credit scoring algorithm governance : the integration of algorithmic regulation and product liability*], 11 Dianzi Zhengwu (电子政务) [E-Government] 28 (2022) (China) (examining using product liability to address issues with credit scoring algorithms and the deficiencies in that approach).

294. *Id.*

295. *Id.*

296. *Id.*

297. *Id.*

298. *Id.*

299. *Id.*

300. *Id.*

301. *See* Jiang Su (江溯), *Zidonghua Juece, Xingshi Sifa yu Suanfa Guizhi—You Lumisi An Yinfa de Sikao* (自动化决策、刑事司法与算法规制———由卢米斯案引发的思考) [*Automated Decision-Making, Criminal Justice, and Algorithm Regulation: Reflections Triggered by the Loomis Case*], 3 Dongfang Faxue (东方法学) [ORIENTAL L.REV.] 76, 86 (2020) (China) (discussing how the integration of automated decision-making systems, such as the COMPAS risk assessment tool in criminal justice, raises critical concerns about algorithmic transparency, accountability, and fairness, necessitating enhanced explainability mechanisms to safeguard due process rights and prevent discriminatory outcomes).

302. *See* Yang Fan (杨帆), *supra* note 293 (explaining that product liability damages may not account for all of the damages incurred from algorithmic credit-scoring).

303. *Id*.

304. *Id*.

reveal, and the law's traditional categories struggle to recognize the full economic and governance significance of score-based harm.[305]

In the EU, product liability is structured around no-fault liability and can, in principle, move the legal focus from proving the producer's fault to assessing whether a product is defective.[306] In its 2024 recast on liability for defective products, the EU extended the material scope of strict liability to software, stating that "[s]oftware is a product for the purposes of applying no-fault liability," regardless of how it is supplied or accessed, including via cloud technologies and software-as-a-service models.[307] On this basis, AI systems can be brought within the product liability architecture.[308] Nevertheless, the move remains contested.[309] The same recitals insist that the Digital Services Act continues to govern intermediary services and that the instrument should not apply to services.[310] Furthermore, despite treating software as a product, the contents of digital files and source code are characterized as information rather than products, exposing borderline cases and system configurations—especially AI as a service and model outputs—to persistent classificatory uncertainty.[311] Recital (2) of Regulation 2022/2065 also frames strict liability as the "sole means" of adequately addressing loss allocations in modern technological production while the recast extends coverage to all types of products, including software.[312]

The recast defines defectiveness by referencing the safety a person is entitled to and pulls foreseeability into the defect inquiry by requiring courts to consider "reasonably foreseeable use" and the "reasonably foreseeable effect" of interconnected products.[313] This structure echoes the older 85/374/EEC framework.[314] Traditionally, the claimant bears the burden of proving the damage, defect, and causal relationship;[315] in AI-related cases, this burden can be structurally difficult to discharge.[316] The recast therefore introduces procedural correctives, including evidence disclosure that is subject to safeguards for trade

305. *Id*.

306. Directive 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products and Repealing Council Directive 85/374/EEC, 2024 O.J. (L 2853) 1, 2 [hereinafter CD2024].

307. *Id.* at 13.

308. *Id.*

309. *See* Ali Enayatollahi & Justine Naessens, *The Revised EU Product Liability Directive: State of Play Across EU Member States and Evolving Risk Landscape*, JONES DAY (June 2026) https://www.jonesday.com/en/insights/2026/06/the-revised-eu-product-liability-directive-state-of-play-across-eu-member-states-and-evolving-risk-landscape#:~:text=Looking%20Ahead%3A%20The%20EU%20PLD%20materially%20reshapes, face%20heightened%20risk%20for%20strict%2C%20no%2Dfault%20claims [https://perma.cc/Z6NW-AEYL] (examining issues with the EU council directive).

310. Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC, 2022 O.J. (L 277) 1, 7–9.

311. *See* CD2024, *supra* note 306, at 13.

312. *Id*. at 2.

313. *Id.* at 30.

314. *Compare* Council Directive 85/374/EEC, 1985 O.J. (L 210) 29, with CD2024, *supra* note 306.

315. *See* David F. Llorca et al., *Liability Regimes in the Age of AI: A Use-Case Driven Analysis of the Burden of Proof*, 76 J. A. I. RSCH. 613, 614–615, 628 (2023) (examining the three traditional methods for victims to obtain compensation for product-induced damages).

316. *Id.* at 628 ("Under product liability, the claimants would have to prove that the AI-based product was defective and the causal link with the damage.").

secrets, and presumptions where technical or scientific complexity would make proof excessively difficult, including where the claimant would otherwise have to explain the "inner workings of an AI system".[317] The allocation of responsibility is further complicated by control along the chain.[318] This is because downstream actors may materially shape outputs through fine-tuning or prompt-based steering in ways that bypass the provider's guardrails.[319] The recast responds to this challenge by widening the set of liable economic operators, such that any person who substantially modifies a product and then makes it available is considered a manufacturer.[320] It also provides for joint and several liability where multiple manufacturers are liable for the same harm.[321] In effect, the instrument reallocates responsibility by reclassifying more participants as manufacturers and subsequently aggregating liability, reducing the practical significance of the traditional model, which concentrated strict liability on a limited set of producers.[322] However, for networked and learning systems where no single actor fully controls performance, multiplying producers and relying on joint and several liability may continue to misspecify control, incentives, and fairness and generate fresh boundary problems between products, services, and information.[323]

## IV. EXPLAINABILITY, RESPONSIBILITY, AND TRADE SECRECY

Before turning to the legal right to algorithmic explanation, it is necessary to address the following questions: Are algorithms at all explainable? Why is the possibility of explanation inseparable from the necessity of responsibility? This

317. CD2024, *supra* note 306, at 48 ("the complex nature of the information and data to be analysed by the claimant; and the complex nature of the causal link, such as a link between a pharmaceutical or food product and the onset of a health condition or a link that, in order to be proven, would require the claimant to explain the inner workings of an AI system.").

318. *See* Llorca et al., *supra* note 315, at 616 (explaining that the complexity of an AI supply chain makes it difficult to decide who is responsible).

319. *Id.*; *cf.* Pengcheng Wen et al., *Not Just the Destination, But the Journey: Reasoning Traces Causally Shape Generalization Behaviors*, in Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, 4-8 (2026) (showing that changing the reasoning content used to fine-tune Qwen3 models produced different patterns of harmful behavior on new tasks, even though the training questions and final answers remained identical).

320. CD2024, *supra* note 306, art 8(2).

321. *Id.* at art. 12(1).

322. *See* Susana Navas, *Producer Liability for AI-Based Technologies in the European Union*, 9 INT'L L. RSCH. 77, 80–82 (2020) (discussing how the liability regime for AI-based technologies in the EU necessitates a reevaluation of traditional product liability concepts, such as "product" and "defect"; "With AI systems there is often more than just one person who 'operates' the technology…the frontend operator is usually the user and the backend operator is the producer, the designer, online service provider or anyone else"; "This backend operator should fall, in the understanding of the Draft Report, under the same liability rules as the producer, manufacturer or developer"); *see also* Jiaming Ji et al., *PKU-SAFERLHF: Towards Multi-Level Safety Alignment for LLMs with Human Preference*, 1 PROC. OF THE 63RD ANNUAL MEETING OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS 31983, 31983–84 (2025) (introducing a large-scale LLM safety-alignment dataset with 44.6k refined prompts, 265k question-answer pairs, safety meta-labels for 19 harm categories and three severity levels, and 166.8k preference data; explaining that safety alignment depends on preference datasets and meta-label classification, and that the authors use the annotation data to train severity-sensitive moderation and safety-centric RLHF algorithms for LLM risk control).

323. *See* Navas, *supra* note 322, at 80.

section starts from a responsibility-based account of control and knowledge and then shifts to a relational account of answerability.[324]

### *A. Are algorithms explainable?*

A right that cannot be operationalized collapses into symbolism rather than accountability.[325] The decisive question is therefore not whether an algorithm can be fully rendered transparent in engineering terms but rather whether it can be explained to a degree that serves the legal functions of justification, contestation, and oversight in the specific decision context.[326] A useful starting point is to resist the rhetorical move that treats opacity as a unique defect of machine systems.[327] Human judgment is a profound black box that cannot be reduced to perfectly rational causal chains, because affect, habit, and social context shape decisions in ways that resist complete articulation.[328] It remains coherent to regulate algorithms more stringently since algorithmic systems can be used by many at once, and thus they replicate the same error at speed and volume, multiplying both the likelihood and the distribution of harm.[329]

The common objection is that only simple models are explainable, while modern machine learning systems, especially deep neural networks, are too complex to interpret even for their designers.[330] While said claim is descriptively plausible in relation to full mechanistic interpretability, it is not decisive for law because legal explanation is not synonymous with reverse engineering.[331] Legal explanation is the practice of understanding and giving reasons in a form appropriate to its audience and purpose—a point that maps cleanly onto algorithmic governance.[332]

From a technical perspective, explainability is understood as an audience-specific design problem, rather than as a single method for opening a black

324. *See infra* Sections IV.A, IV.B (answering the preceding questions).

325. *See* Kaminski, *supra* note 36, at 213 (discussing the relationship between transparency and the exercise of other substantive rights).

326. *See* Michele Loi et al., *Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms*, 23 ETHICS & INFO. TECH. 253, 253–54 (2021) (discussing how transparency as design publicity, which involves explaining algorithms by their intended goals, translation into measurable criteria, and performance metrics, can enhance the explainability and justification of algorithmic decision; "transparency. . .is valuable because and in so far as it enables individuals. . . to assess whether these decisions are morally and politically justifiable," and post-hoc explanations alone "fail to explain why a decision. . .is normatively adequate.").

327. *See infra* text accompanying notes 328–32.

328. *See* KENNETH R. HAMMOND, HUMAN JUDGMENT AND SOCIAL POLICY: IRREDUCIBLE UNCERTAINTY, INEVITABLE ERROR, UNAVOIDABLE INJUSTICE 316–17 (Oxford Univ. Press, 2000) (discussing irreducible uncertainty in human judgment, the influence of contextual and framing effects, and the limits of articulating decision rationales).

329. *See* Wu, *supra* note 271, at 124–25 (discussing how black-box algorithms cause disparate harms to a wide variety of people).

330. *Id*.; Philipp Hacker et al., *Explainable AI Under Contract and Tort Law: Legal Incentives and Technical Challenges*, 28 A.I. & L. 415, 430 (2020).

331. *See generally* Boris Babic & I. G. Cohen, *The Algorithmic Explainability "Bait and Switch"*, 108 MINN. L. R. 857 (2023) (discussing the concepts of "explainable AI" and "interpretable AI" insofar as they serve useful legal distinctions).

332. *See* Ashley Deeks, *The Judicial Demand For Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1842–47 (2019) (discussing how legal explanations change based on the legal audience).

box.[333] Arrieta and coauthors define explainable AI by reference to an audience, namely, a system that produces details or reasons that make its functioning clear or easy to understand for the relevant recipient.[334] On that view, explanation depends on who needs the account, what task the account serves, and how the explanation will be evaluated.[335] A range of methods now serves this explanatory function, including feature-attribution tools, local surrogate explanations, rule-based explanations, example-based explanations, visual explanations, and counterfactual or recourse-oriented explanations.[336] In credit scoring, explanation methods are often differentiated by function.[337] Global explanations support model-level understanding, local feature-based explanations identify the factors relevant to an individual prediction, and local instance-based explanations compare the applicant with similar cases.[338] Recent XAI work reinforces the same point by separating technical, human-centered, and societal explanation layers, with natural-language mediation used to translate technical outputs into explanations that different stakeholders can actually use.[339]

These techniques can be valuable, but their value is bounded by technical and institutional limits.[340] Local *post hoc* methods, such as LIME, approximate a model's behavior near a particular decision, while feature-attribution methods, such as SHAP, assign importance values to input features for a prediction.[341] In a credit case, such methods might show that the debt-to-income ratio, recent delinquencies, utilization rate, or income volatility contributed to the denial, but the output remains an approximation rather than a full reconstruction of the

333. *See* Babic & Cohen, *supra* note 331, at 882 (explains how explainable AI does not necessarily require opening the black box).

334. *See* Alejandro B. Arrieta et al., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI*, 58 INFO. FUSION 82, 84–86 (2020) (defining XAI by reference to explanations produced for an audience and surveying XAI concepts and taxonomies).

335. *See* Finale Doshi-Velez & Been Kim, *Towards a Rigorous Science of Interpretable Machine Learning,* ARXIV MACH. LEARNING, at 2–5 (2017) (arguing that interpretability depends on the task, audience, and evaluation setting).

336. *See* Marco Tulio Ribeiro et al., "*Why Should I Trust You?" Explaining the Predictions of Any Classifier*, PROC. 22D ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING, Aug. 2016, at 1135–40 (introducing local interpretable model-agnostic explanations for individual predictions) [hereinafter Ribeiro]; *see also* Scott M. Lundberg & Su-In Lee, *A Unified Approach to Interpreting Model Predictions*, 30 PROC. 31ST INT'L CONF. ADVANCES NEURAL INFO. PROC. SYS. (2017) (developing SHAP as a unified feature-attribution framework).

337. *See* Wu, *supra* note 271, at 72–76 (contextualizing the landscape of automated credit scoring).

338. *See* Lara M. Demajo et al., *An Explanation Framework for Interpretable Credit Scoring*, 12 INT'L J. A.I. & APPLICATIONS 19, 24–29 (2021) (distinguishing global, local feature-based, and local instance-based explanations for credit scoring) [hereinafter Lara].

339. *See* Marilyn Bello et al., *A Three-Level Framework for LLM-Enhanced eXplainable AI: From Technical Explanations to Natural Language*, INFO. SYS. FRONTIERS at 12–22 (2025) (proposing a multi-level framework that aligns technical, human-centered, and societal explanations with different stakeholders).

340. *Id.* at 16; *see also* Doshi-Velez & Kim, *supra* note 335, at 9 (expressing the institutional limits of classification of such models); *cf.* Chuxue Cao et al., *Pushing the Boundaries of Natural Reasoning: Interleaved Bonus from Formal-Logic Verification in Language Models*, 306 PROC. MACH. LEARNING RSCH. 11641, 11653 (2026), https://proceedings.mlr.press/v306/cao26y.html [https://perma.cc/Z2MQ-PT3S] (acknowledging that ambiguous or commonsense-heavy natural-language descriptions may introduce mapping errors during formalization and produce incorrect verification feedback).

341. *See* Ribeiro, *supra* note 336, at 1135–38 (explaining how LIME works as a *post hoc* method); Lundberg & Lee, *supra* note 336, at 4765–68 (explaining how SHAP functions as a feature-attribution model).

model's reasoning.[342] Rudin's critique is especially important for high-stakes credit decisions because *post hoc* explanation should not be treated as a complete substitute for an inherently interpretable model where one can achieve comparable performance.[343] Slack and coauthors show a sharper technical flaw: LIME and SHAP can be adversarially manipulated so that a biased classifier appears innocuous to the explanation method.[344] A broader methodological critique adds that XAI still lacks sufficiently formal notions of explanation correctness, and that popular feature-attribution methods may fail when feature dependence and the data-generating structure are not properly taken into account.[345]

A more workable approach is therefore to distinguish between internal and external explanations, since they answer different questions and are assessed by different criteria.[346] The internal form is developer-facing and aimed at diagnostics, robustness, and liability allocation, which aligns with the concern that systems can behave in ways that lack a clear principle.[347] Core materials include validation records, data-lineage documentation, subgroup-performance analysis, and records of known model limits.[348] External explanations are user- and regulator-facing and aimed at legal intelligibility and public accountability, rather than at model transparency as such.[349] For the affected borrower, the relevant account should identify the principal reasons for the adverse outcome, the data categories involved, the role of automated scoring, any correctable information,

342. *See* Ribeiro, *supra* note 336, at 1136–39 (explaining how local explanations support trust and model diagnosis); *see also* Doshi-Velez & Kim, *supra* note 335, at 5–8 (distinguishing evaluation settings for interpretability).

343. *See* Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206, 206–15 (2019) (explaining key issues with explainable ML models).

344. *See* Dylan Slack et al., *Fooling LIME and SHAP: Adversarial Attacks on Post Hoc Explanation Methods*, AAAI/ACM CONF. ON AI, ETHICS, & SOC'Y, 2020, at 181–85 (showing that post hoc explanation methods can be manipulated to conceal discriminatory behavior).

345. *See generally* Stefan Haufe et al., *Explainable AI Needs Formalization*, 2 NPJ A.I., 2026 (arguing that XAI evaluation often lacks formal explanation correctness and that feature-attribution methods can fail under feature dependence).

346. Maja Brkan & Gregory Bonnet, *Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas*, 11 EURO. J. RISK REGUL. 18, 33 (2020); *see also* Arrieta et al., *supra* note 334 (providing taxonomies and definitions for the various processes of XAI).

347. *See* David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 123–26, 141–43 (2014) (discussing how liability rules for autonomous thinking machines, such as driverless cars, necessitate rethinking traditional product liability frameworks to address challenges in tracing and attributing accidents to human error or machine design).

348. *See* Doshi-Velez & Kim, *supra* note 335, at 5–8 (discussing data documentation for problems in interpretability and explainability); *see also* Ribeiro, *supra* note 336, at 1136–39 (provides explanation and application for local model limits); Lara, *supra* note 338, at 24–29.

349. *See* J Tang Jun (唐钧), *Zuihou de Anquan Wang—-Zhongguo Chengshi Jumin Zuidi Shenghuo Baozhang Zhidu de Kuangjia* (最后的安全网——中国城市居民最低生活保障制度的框架) [*The Final Safety Net: Framework of Minimum Livelihood Security for Urban Residents*], 1 ZHONGGUO SHEHUI KEXUE (中国社会科学) [SOC.SCI.CHINA] 117, 119–21 (1998) (discussing how the establishment of an urban residents' minimum living security system in China addresses the increasing poverty issue among urban residents, providing a safety net that considers the limitations of traditional social assistance systems and emphasizing the need for transparency and fairness in the allocation of resources).

and the path to human review or further assessment.[350] Once this distinction is made, the argument that the public cannot understand technical disclosures ceases to be a refutation and instead becomes an institutional design constraint.[351] It is true that source code or high dimensional parameters are not legible to most recipients, and even perfect disclosure can fail to equalize power if only the controller can meaningfully interpret what is disclosed.[352] Nevertheless, this point supports—rather than undermines—the case for external explanation duties, because law regularly demands intelligible reasons rather than technical artifacts.[353] Therefore, it is a category mistake to reduce a right to explanation to code access, as doing so ignores its normative role in securing procedural fairness in high-stakes decisions, such as credit denials.[354]

Counterfactual explanations occupy a middle position between technical diagnosis and legal contestation.[355] In analytic terms, a counterfactual asks what would have happened if some fact had been different: if variables V had values v rather than v', then score p rather than p' would have been returned.[356] The explanation might state that the application would have been approved if the verified income had been higher, the outstanding debt had been lower, or the disputed delinquency had been corrected, while all other relevant variables remained constant.[357] Counterfactual explanation is attractive for practical reasons because it can give the applicant a usable account of what would have changed the outcome without disclosing the entire model or proprietary scoring logic.[358] More recent work has tried to make recourse less individualized and more useful at scale.[359] Kavouras and coauthors define global counterfactual explanations as a small set of actions that should be interpretable, low-cost, and effective for as many affected individuals as possible.[360] Counterfactuals also have limits that

350. *See* Kaminski, *supra* note 36, at 193–99 (discussing explanation as a governance tool for accountability, contestation, and oversight); *see* Edwards & Veale, *supra* note 37, at 55–64 (warning that explanation rights must be tied to contestability and institutional function rather than abstract transparency).

351. *See generally* Rudin, *supra* note 343 (discussing and criticizing the institutional constraints of explainable models).

352. *See* Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 640–46 (2020) (arguing that limits on public understanding of technical disclosures do not defeat transparency claims but instead operate as institutional design constraints, and that disclosure alone may fail to equalize power where interpretive capacity remains asymmetric).

353. *See* Babic & Cohen, *supra* note 331, at 904–09 (discusses the applicability of interpretable AI to procedural justice in lieu of explainable AI).

354. *See* Wu, *supra* note 271, at 115–20 (discussing the failures of pure algorithmic scrutiny in credit scoring).

355. *See* Bello et al., *supra* note 339, at 16–18 (discusses counterfactuals as a useful tool for the mid-level approach within their proposed framework).

356. *See* Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. & TECH. 841, 850–54 (2018) (defining counterfactual explanations by reference to changes in input variables and contrasting possible outcomes); *see* also DAVID LEWIS, COUNTERFACTUALS, at 1–4, 84–91 (1973); JUDEA PEARL, CAUSALITY, 223–24 (Cambridge Univ. Press, 2009).

357. *See* Wachter et al., *supra* note 356, at 879–84 (arguing that counterfactual explanations can support understanding, contestation, and future action without opening the black box).

358. *Id.* at 845–48, 879–84.

359. *See* Kavouras et al., *infra* note 360.

360. *See* Loukas Kavouras et al., *GLANCE: Global Actions in a Nutshell for Counterfactual Explainability*, 40 PROC. AAAI CONF. ON A.I. 22545, 22545–53 (2026) (defining global counterfactual explanations for the purposes of AI models).

matter for law.[361] A single counterfactual may be too narrow if the suggested change is impossible, unstable, too costly, or based on information unavailable to the borrower.[362] Reason codes and multiple counterfactuals may sometimes match or outperform single counterfactual explanations, especially where unknown real-world constraints affect whether a person can successfully reapply.[363]

The question is therefore not whether XAI can fully reveal an algorithm, but how the law should require that reasons be presented in a form that affected persons and reviewing institutions can understand, evaluate, and contest.[364] This is consistent with the jurisprudential insight that explanation involves both understanding and a proper form of outward presentation, since justification is not complete until it is communicated in a manner that can be evaluated by others.[365] Counterfactual explanations, often discussed in the European context, illustrate both the promise and the limits of this approach.[366] The Court of Justice of the EU (CJEU) has indicated that an intelligible explanation may be delivered in counterfactual form because it can show how different inputs could have produced a different outcome, while also making it clear that transparency is not achieved by technical data dumping, since "*[t]hose requirements cannot be satisfied either by the mere communication of a complex mathematical formula, such as an algorithm, or by the detailed description of all the steps in automated decision-making, since none of those would constitute a sufficiently concise and intelligible explanation.*"[367]In this spirit, one strand of scholarship argues that counterfactual explanations align with the GDPR's contestability function by translating decision logic into a practically usable account of what would have changed the result.[368]

One critique argues that this promise can collapse in discrimination-sensitive settings.[369] Counterfactuals tend to highlight only adjustable inputs as levers for change while leaving identity linked and effectively immutable drivers either

361. *See* Wachter et al., *supra* note 356, at 863–65 (discussing counterfactual explanations and their legal limits).

362. *See* Rudin, *supra* note 343, at 209–11 (criticizing counterfactual explanations as insufficient for high-stakes black-box decisions).

363. *See generally* Sohini Upadhyay et al., *Counterfactual Explanations May Not Be the Best Algorithmic Recourse Approach*, PROC. 30TH INT'L. CONF. ON INTELLIGENT USER INTERFACES 446 (2025) (explaining reason codes and multiple counterfactuals as potential alternatives to single counterfactuals).

364. *See generally* Babic & Cohen, *supra* note 331 (discussing the potential legal standards for explainable and interpretable AI).

365. *See* Loi et al., *supra* note 326, at 260 (describing the standards of explainability and legibility consistent with past jurisprudence).

366. *See* Wachter et al., *supra* note 356, at 863–65 (discussing counterfactual explanations in the context of the GDPR's right to explanation).

367. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶ 59.

368. Wachter et al., *supra* note 88, at 4 ("Data subjects are granted additional safeguards to obtain human intervention, express views, or contest a decision . . . but *not* to obtain an explanation of the decision reached." (emphasis added)).

369. *See infra* note 371, at 84 (discussing how counterfactuals do not always best address or respond to concerns with discrimination).

undisclosed or treated as beyond contestation;[370] thus, they offer an appearance of agency without revealing whether the decision rests on impermissible or structurally unfair proxies.[371] This brings a further governance risk into view, namely the risk of selective legibility, where institutions prioritize what is easiest for a layperson to grasp, and controllers comply through clear yet incomplete curated narratives, which leaves problematic modeling choices insulated from scrutiny.[372] Arguably, this becomes a pyrrhic victory for data subjects because clarity can be purchased at the price of depth.[373] It follows that the legal problem is not that lay people cannot read code, but that without a layered disclosure model, the controller can monopolize interpretive authority and convert explanation into public relations.[374] This is an obvious conflict of interest in any system where the controller also controls the evidential record that would make contestation effective.[375]

Chinese practice also suggests that outcome-focused accountability can operate without code disclosure.[376] In *Wei Zexi*, regulators pressed for changes to the credibility weighting logic rather than demanding access to source code, which indicates that regulatory leverage can attach to decision logic and its effects.[377] Similarly, in *Ren v. Baidu*, the Court accepted Baidu's account of its autocomplete function without compelling a disclosure of internal code, which signals that courts may be satisfied when an operator provides a credible functional narrative that can be assessed against observable outputs.[378] This approach

370. *See id.* (highlighting the implications of people not fully understanding their inputs and not having controls over the underlying factors that drive automated decision-making).

371. *See* Solon Barocas et al., *The Hidden Assumptions Behind Counterfactual Explanations and Principal Reasons*, PROC. OF THE 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 80, 84 (2020) (discussing how the hidden assumptions behind counterfactual explanations and principal reasons in algorithmic decision-making can mislead decision subjects by obscuring the true factors influencing decisions, highlighting the need for more nuanced and transparent approaches to ensure explainability and fairness).

372. *Id.* at 80 (discussing the "unchecked power" that feature-highlighting explanations grant decision makers and the subjective choices involved in determining which features to present); Baoyang Chen, *Selective Legibility: From Material Readout to Machine Agency*, 7 INNOVATION 1, 2 (2026) (discussing selective legibility).

373. *See id.* (explaining that the simplicity of feature-highlighting explanations comes from selectively presenting information, which may leave decision subjects with an incomplete account of the decision); Chen, *supra* note 372, at 2 ("Selective legibility is partial by design. Every readout selects, and every interface narrows. The task is to design that narrowing honestly, orienting people without pretending to settle the matter and leaving room for judgment.").

374. Huq, *supra* note 352, at 640–41.

375. *See* Wachter et al., *supra* note 88, at 21 (discussing the inherent tension between controllers and data subjects in the effectuation of data subject rights and requests regarding automated decision making).

376. *See China to Regulate Search Results Following Man's Death*, ASSOCIATED PRESS (June 25, 2016, 1:02 AM), https://apnews.com/general-news-international-news-e77086356c094b9aa426bfc89ed0ed10 [https://perma.cc/JF8M-79GN] (explaining how Chinese regulators responded to regulate Baidu after the death of Wei Zexi).

377. *See id.*; *see also* Ian Lau, *Baidu Advertising Changes After the Wei Zexi Scandal*, DRAGON METRICS, (June 29, 2016) https://www.dragonmetrics.com/baidu-advertising-after-wei-zexi-scandal/ [https://perma.cc/KF77-KU2F] (highlighting the specific changes made by Baidu in response to government regulation).

378. *See* Chen Changyi (陈昶屹), *Xianyou Falu Tixi xia "Bei Yiwang Quan" Anjian de Shenli Silu ji Baohu Lujing —- Cong Woguo "Bei Yiwang Quan" Di Yi An Shuoqi* (现有法律体系下"被遗忘权"案件的审理思路及保护路径———从我国"被遗忘权"第一案说起) [*Trial Ideas and Protection Paths for "Right to be Forgotten" Cases Under the Existing Legal System—Starting from China's First "Right to be Forgotten" Case*], 2 FALU SHIYONG (法律适用) [L.APPL.] 41, 41–43 (2017)

is not automatically pro-defendant because outcome-based verification can still expose discriminatory patterns and inconsistencies while respecting trade secrets and security.[379] Therefore, most decision systems are explainable in layered and purpose-sensitive ways, even when they are not fully interpretable at the level of code or parameters.[380] Existing explainable AI initiatives often prioritize technical interpretability over normative accountability, and firms may even manipulate explanations to serve business objectives rather than fairness, which is precisely why legal standards must specify both content and verifiability.[381]

### *B. Responsibility in automated credit governance*

Once a system materially shapes another person's options, the demand for justification must track the conditions under which responsibility can be attributed.[382] The classical responsibility template requires control and knowledge.[383] The agent should be capable of controlling the action and know what is being done and with what object, such that coercion and ignorance cannot defeat responsibility.[384] The law still requires someone to answer;[385] however, the evidential and epistemic structure of algorithmic systems makes it unreasonable to require the affected party to prove what the controller can

(discussing how the "right to be forgotten" in the EU, as established through judicial rulings, raises questions about its applicability and legal basis in China, emphasizing the need for a careful balance between personal information protection and public interests such as freedom of speech and information flow); *see also* Ren Moumou v. Baidu Online Network Technology (Beijing) Co., Ltd. (任某某诉北京百度网讯科技有限公司侵害一般人格权纠纷案), Beijing Haidian Dist. People's Ct. (2015) Hai Min Chu Zi No. 17417, aff'd, Beijing First Intermediate People's Ct. (2015) Yi Zhong Min Zhong Zi No. 09558 (China) (accepting the platform's functional explanation of its autocomplete mechanism based on observable outputs and declining to require disclosure of internal algorithms or source code).

379. *See* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 174 (2018) (explaining that post-implementation validation can identify "output biases" and that confidential third-party audits can provide accountability where disclosure is limited by trade-secret, security, or privacy concerns).

380. *See* Margot E. Kaminski & Gianclaudio Malgieri, *Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations*, 11 INT'L DATA PRIV. L. 125, 141–43 (2021) (proposing multi-layered explanations that provide system-level, group-level, and individualized information about algorithmic decision-making).

381. *See* Walter A. Mostowy, *Explaining Opaque AI Decisions, Legally*, 35 BERKELEY TECH. L. J. 1291, 1303 (2020) (citing Lisa Käde & Stephanie von Maltzan, *Towards a Demystification of the Black Box—Explainable AI and Legal Ramifications*, 2 J. INTERNET L. 3, 5 (2019)) (explaining the logic behind the model's decision process as a whole in a general context involves providing a global explanation).

382. *See generally* Coeckelbergh, *infra* note 386 (discussing the issues with responsibility attribution raised by the increased use of AI technologies and Aristotelian theories on voluntary acts and conditions for responsibility).

383. *See id.* at 2052 (discussing the traditional governance framework based on responsibility and control of AI technologies).

384. *See* ARISTOTLE, NICOMACHEAN ETHICS 39–40 (David Ross trans., Oxford Univ. Press, 2010) (distinguishing voluntary action from acts performed under compulsion or ignorance).

385. GDPR, art.VII, 2016 O.J. (L 119) 40.

reveal.[386] The normative pivot is therefore from causation to answerability.[387] Responsibility not only encompasses the cause of harm but also providing a reasoned response to those subject to the system that can be questioned, corrected, and acted upon.[388] This stance aligns with the autonomy-based view of persons as agents who are entitled to rational engagement rather than passive recipients of opaque classification.[389] Accordingly, the right to algorithmic explanation is not a demand that the machine must possess moral agency but that the human institutions organizing the socio-technical network provide intelligible, contestable grounds, because power is exercised through that network and cannot be legitimized by hiding behind technical opacity.[390]

This is where Kant's maxim that "ought implies can"[391] can be effective, but only if it is applied to the right targets and levels of abstraction.[392] Rather than claiming that an algorithm, as such, has moral duties, the maxim is invoked to ground a normative claim regarding the institutional choice to deploy algorithmic inference in rights-affecting contexts.[393] If a controller can scale causally efficacious inferences across millions of people, it can also scale the complementary capacities making those inferences answerable: interpretable justification, contestation pathways, and auditable traces.[394] By contrast, if the controller claims that the model cannot be explained in causally grounded terms, the said

386. *See* Mark Coeckelbergh, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, 26 SCI. & ENG'G ETHICS 2051, 2064 (2019) (discussing how responsibility attribution in AI systems necessitates a relational approach that considers both the moral agents using AI and the moral patients affected by its decisions as users "rightly demand reasons and explanations for actions and decisions [made by using AI] that affect them").

387. *See id.* (discussing the pivot to demanding answerability from AI agents and systems, demonstrated by AI giving reasons and explanations for how it made its decisions); *see* Rebecca Williams et al., *From Transparency to Accountability of Intelligent Systems: Moving Beyond Aspirations*, 4 DATA & POL'Y 7, 20 (2022) (explaining that accountability is fundamentally a matter of "answerability," or the need to explain or give an account).

388. *Id.*

389. *See generally* Sarah Buss & Andrea Westlund, *Personal Autonomy*, STAN. ENCYCLOPEDIA PHIL. (2018), https://plato.stanford.edu/archives/fall2018/entries/personal-autonomy/ [https://perma.cc/WC8R-3W3C] (discussing how ensuring personal autonomy in algorithmic decision-making processes necessitates a nuanced understanding of the conditions under which individuals can be considered to govern their own actions).

390. *See* BRUNO LATOUR, WE HAVE NEVER BEEN MODERN 64–65, 89–90 (Catherine Porter trans., Harv. Univ. Press, 2012) (discussing how "[t]here is nothing of the sort when technological reason has to be kept as remote as possible from the free discussion of human beings" and attributing action to objects is not to absolve humans of responsibility).

391. IMMANUEL KANT, RELIGION WITHIN THE BOUNDARIES OF MERE REASON 70 (Cambridge Univ. Press) ("For if the moral law commands that we *ought* to be better human beings now, it inescapably follows that we must be *capable* of being better human beings.").

392. *See* Yaffe, *infra* note 393, at 219 (discussing the how "ought implies can" maxim does not apply to all circumstances and algorithms).

393. *See* Gideon Yaffe, *'Ought' Implies 'Can' and the Principle of Alternate Possibilities*, 59 ANALYSIS 218, 220–21 (1999) (discussing how reconciling the "Ought Implies Can" maxim with the rejection of the Principle of Alternate Possibilities (PAP) is crucial for maintaining the moral coherence and explainability of algorithmic decision-making frameworks, as the maxim asserts that moral obligations must align with feasible actions).

394. *See* Joris Hulstijn, *Computational Accountability*, PROC. OF THE NINETEENTH INT'L CONF. ON A.I. AND L., 121, 126–27 (2023) (discussing how integrating explainability into automated decision-making systems is essential for ensuring computational accountability, as it requires transparent justification of decisions, enabling affected parties and auditors to understand and challenge the rationale behind algorithmic outcomes, thereby fostering trust and adherence to ethical guidelines).

claim does not neutralize the duty;[395] rather, it indicts the appropriateness of using the model for decisions that restructure entitlements and opportunities.[396] In other words, a system that can sincerely produce an outcome can—and therefore should—be held to a standard of sincerity in its reasons.[397] Algorithmic cognition, like human cognition in Kant's architecture, is enabled and bounded by its organizing forms, regardless of whether they are categories or model architectures.[398] Its outputs are thus constrained by what its data and objective functions can express, and there lies a risk not merely of error but of creating an illusion of necessity when a bounded inferential system is treated as if it disclosed the thing itself rather than a constructed judgment within its own parameters.[399] The practical upside is that explanation must reveal those organizing constraints and their domains;[400] otherwise, the system invites a form of antinomy in adjudication and governance.[401] Thus, two apparently rigorous inferences can yield opposite conclusions and, lacking interpretable grounds, neither can be challenged as more than a claim over authority.[402] The legal implication is that explanation should be designed as an institutional practice rather than a technical afterthought.[403] This includes placing a genuinely effective human in the loop with obligations of feedback and revision and equipping affected persons with operational powers to know, question, and contest a direction.[404]

395. *See* GDPR, art. 22(3), 2016 O.J. (L 119) 1, 46 (requiring controllers to implement safeguards for qualifying automated decisions, including human intervention and opportunities for data subjects to express their views and contest decisions).

396. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1133–35 (2018) (explaining that the inability to provide adequate reasons for consequential automated decisions raises questions about whether such systems are appropriate for decisions requiring justification).

397. *See* Qing Lyu et al., *Towards Faithful Model Explanation in NLP: A Survey*, 50 COMPUTATIONAL LINGUISTICS 657, 660–61 (2024) (explaining that a faithful explanation must accurately reflect the reasoning process underlying a model's prediction rather than merely provide a plausible account).

398. *See* Hulstijn, *supra* note 394, at 125 (describing how computer and AI logic is based on the underlying logic and systems used to build it).

399. *See* Andrew D. Selbst et al., *Fairness and Abstraction in Sociotechnical Systems,* PROC. CONF. FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 59, 59–61 (2019) (explaining that machine-learning systems necessarily abstract from social context and warning that treating a bounded technical model as representing the entire relevant system can produce inaccurate or misguided conclusions); *cf.* Chengyi Ju et al., *Benchmarking Multi-National Value Alignment for Large Language Models*, *in* FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: ACL 2025, 20042, 20049–50 (Wanxiang Che et al. eds., 2025), https://doi.org/10.18653/v1/2025.findings-acl.1028 (reporting improved alignment after preference optimization on news-derived value statements and acknowledging that those datasets may not fully represent a country's values or its population's diverse perspectives).

400. Margaret Mitchell et al., *Model Cards for Model Reporting*, PROC. CONF. FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 220, 220–21 (2019) (proposing model documentation that identifies intended uses, relevant application domains and conditions, performance characteristics, and limitations).

401. *Id.*

402. *See* JOHN DANAHER, AUTOMATION AND UTOPIA: HUMAN FLOURISHING IN A WORLD WITHOUT WORK (Harv. Univ. Press, 2019).

403. *See* Mostowy, *supra* note 381, at 1323 (discussing the inherent requirements to disclose explanations of automated decision-making based on principles derived from EU principles on human rights).

404. *See* Kevin Baum et al., *From Responsibility to Reason-Giving Explainable Artificial Intelligence*, 35 PHIL. & TECH. 12, 3–7 (2022) (discussing how providing reason-giving explainability in AI systems is crucial for closing moral responsibility gaps, as it enables human decision-makers to understand and evaluate the motivations behind AI recommendations, ensuring that decisions are made with sufficient epistemic access and moral accountability).

### *C. Do explanation rights conflict with trade secrets?*

The GDPR recognizes that rights to access and explanation operate within limits set by the rights and freedoms of others, including trade secrets and intellectual property;[405] however, it also confines any restriction to what is necessary and proportionate and requires that the essence of the right be respected.[406] The question is not whether explanation rights should override trade secrets but rather what content is required for contestation and which institutional channels can it be delivered by without collapsing the controller's position.[407]

*Dun & Bradstreet Austria* is pivotal because it specifies the legally required depth of explanation in a manner that narrows the surface of conflict with trade secrets.[408] The Court treated multilingual divergence as legally informative when construing Article 15(1)(h), noting that the English word "meaningful" has different equivalents in other languages, concluding that these senses are complementary rather than competing.[409] It then interpreted "logic involved" by reference to, *inter alia*, Czech and Polish versions translating the phrase as "procedures" and "principles";[410] thus, it concluded the "logic involved" in Article 15(1)(h) covers all relevant information concerning the procedure and the principles of a specific result.[411] Simultaneously, trade secrets cannot justify an absolute refusal of information; when controllers invoke Article 2(1) of the Trade Secrets Directive (EU) 2016/943, they must submit the allegedly protected materials to a supervisory authority, which then conducts case-specific balancing, as opposed to blanket exclusions.[412] Therefore, *Dun & Bradstreet* recasts meaningful explanation as the disclosure of procedures and principles, subject to institutional review, substantially reducing the need to open the full black box to data subjects.[413]

Building on this argument, the apparent conflict is further mitigated if explanation duties are designed in a layered and audience-specific manner reflective of how different actors actually use explanations in credit scoring.[414] Empirical and technical studies suggest that loan officers prefer local sample-based explanations that compare a rejected applicant's profile to similar successful or

405. GDPR, 2016 O.J. (L 119) 12; GDPR, art.XV, 2016 O.J. (L 119) 43.

406. GDPR art. XV, 2016 O.J. (L 119) 43; GDPR, art. 23(1) 2016 O.J. (L 119) 46–7.

407. *See* Gianclaudio Malgieri & Giovanni Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 243, 254–55 (2017) (arguing that meaningful explanation can be reconciled with trade-secret protection by calibrating the scope of disclosure and the institutional recipient of protected information).

408. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 58–59 (holding that the providing "meaningful information about the logic involved" requires the explanation to be provided in a "concise, transparent, intelligible and easily accessible form").

409. *Id.* ¶ 41 ("The diversity of generally accepted meanings in the various language versions must be understood in such a way that the various meanings set out in the preceding paragraph are complementary.").

410. *Id.* ¶ 42.

411. *Id.*

412. *Id.* ¶ 76.

413. *Id.* ¶¶ 61, 66, 76; *see* Wachter et al., *supra* note 356, at 4 (discussing how to provide information to data subjects without disclosing the "black box" of algorithmic decisions subject to trade secrets or intellectual property).

414. Lara, *supra* note 338, at 19–20.

unsuccessful cases;[415] rejected applicants prefer local feature-based explanations that identify the specific factors that caused rejection;[416] and regulators or data scientists prefer global model explanations.[417] The EU AI Act reinforces this layered architecture by classifying credit scoring as a high-risk use case.[418] Under Article 6(2), AI systems listed in Annex III are treated as high risk, and Annex III section 5(b) covers "AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score."[419] The providers must retain detailed technical documentation for national authorities under Article 11 and Annex IV, whereas deployers and end users must receive "concise, complete, correct and clear information that is relevant, accessible and comprehensible" under Article 13.[420] In practice, this means that the most sensitive elements of the model architecture and training data can remain within a confidential regulatory perimeter while individuals obtain sufficiently concrete information to support their rights of contestation under Articles 15 and 22.[421] As a result, a layered explanation substantially deflates the claim that explanation rights lead to the destruction of trade secret value.[422]

Additionally, in *SCHUFA*, national courts had previously treated the protection of the trade secret (i.e., the scoring formula) as a reason to limit individual scrutiny.[423] However, the subsequent reference prompted Advocate General Pikamäe to state that although protecting trade secrets or intellectual property is a legitimate reason to refuse the disclosure of the scoring algorithm, it cannot justify an absolute refusal, particularly where appropriate forms of communication exist that can facilitate understanding while preserving confidentiality.[424] Chinese practice points to the same direction.[425] In *Chen v Hangzhou Software Service Company*, the Court recognized the platform's filtering algorithm as a trade secret but required a neutral expert panel review and obliged the platform to provide a reasonable explanation of the general principles and concrete operation of the filtering system to the claimant.[426] These cases indicate that trade secrets are themselves subject to legal limits when they threaten to hollow out

415. *See id.* at 24–25 ("There are three main personas that require explainability in credit scoring, being; (i) loan officers that are said to prefer local sample-based explanations, (ii) rejected loan applicants that are said to prefer local feature-based explanations, and (iii) regulators or data scientists that are said to prefer global model explanations").

416. *Id.*

417. *Id.*

418. EU AI Act, at 1 ("AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud").

419. *Id.*

420. *Id.* ("High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to deployers").

421. *Id.*

422. *Id.*

423. Case C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, ¶ 16.

424. *Id.* ¶¶ 43–46.

425. Chen v. Hangzhou Alimama Software Service Co., HANGZHOU RAILWAY TRANSP. CT., (2017) Zhe 8601 Min Chu No. 3306 (China).

426. *Id.*

fundamental rights.[427] A normatively defensible settlement is therefore a calibrated transparency regime that insists on channels for deeper disclosures through independent bodies and treats trade secrets as a factor in proportionality analysis, as opposed to treating them as a trump card against algorithmic accountability.[428]

## V. FROM NOTICE TO CONTESTABILITY

### *A. Is the right of access or to be informed adequate?*

A credible claim that an algorithmic explanation right exists should begin by demonstrating why the rights of access and to be informed, even when formally satisfied, do not deliver the governance functions required by automated credit scoring.[429] While the traditional notice-and-consent model presupposes a stable object of consent, automated credit scoring represents a moving target across the data collection, model design, deployment, and iterative optimization stages.[430] Thus, users often only receive abstract risk prompts rather than an intelligible account of an evolving decision logic.[431] A purely *ex ante* model treats users as having authorized unknown future processing risks, turning consent into a liability waiver rather than an autonomy-enabling device.[432]

Given that consent rules supply legality but not contestability, Chinese law illustrates the indispensability and limits of informed consent.[433] Article 1035(1) of the Civil Code and Article 14 of the PIPL establish informed consent as the legal basis for lawful processing, because they aim to ensure that individuals can express their intent based on sufficient information.[434] However, many user agreements satisfy this duty through formalistic prompts (e.g., statements) and the practice strengthens platform dominance by confining users to surface-level information.[435] Even if one assumes clearer drafting, an exhaustive *ex ante*

427. Case C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, ¶¶ 43–46; Chen, *supra* note 425.

428. *SCHUFA Holding (Scoring)*, ECLI:EU:C:2023:957, ¶¶ 43–46; *Chen*, Zhe 8601 Min Chu No. 3306; Loi et al., *supra* note 326, at 258; *see also* Citron & Pasquale, *supra* note 24, at 21 ("But in general, we need to switch the default in situations like this away from an assumption of secrecy, and toward the expectation that people deserve to know how they are rated and ranked").

429. *See* Wu, *supra* note 271, at 115–20 (discussing the failures of pure algorithmic scrutiny in credit scoring); *see also* Citron & Pasquale, *supra* note 24, at 21 ("Individuals should have the right to inspect, correct, and dispute inaccurate data, and to know the sources (furnishers) of the data. Ironically, some data brokers now refuse to give out their data sources because of 'confidentiality agreements' with sources.").

430. Antonio Carrascosa, *AI, Credit Risk Management and Data Protection*, DO BETTER (Dec. 19, 2024), https://dobetter.esade.edu/en/AI-banks-credit [https://perma.cc/Q8Q2-9SXM]; *see also* Citron & Pasquale, *supra* note 24, at 27–30 (discusses the problems with notice in automated scoring).

431. *See* Citron & Pasquale, *supra* note 24, at 5 ("Although some scores, such as credit, are available to the public, the scorers refuse to reveal the method and logic of their predictive systems. No one can challenge the process of scoring and the results because the algorithms are zealously guarded trade secrets.").

432. *See* Bagheri & Nakajima, *infra* note 436, at 10–12 (discusses the allocation of risk in an *ex ante* model).

433. *See, e.g.*, Chinese Civil Code, art. 1035(1); *see also* PIPL, art. 14.

434. Cheng Xiao (程啸), *supra* note 208; PIPL, art. 14.

435. *See, e.g.*, Tencent, *supra* note 23 (providing a WeChat pay score user service).

explanation cannot solve the core problem.[436] As a static term cannot anticipate how a model will evolve, later harm will be reframed as an *ex post* unforeseeable responsibility rather than a governance failure.[437] The *Zhima Credit* arrangement illustrates this structural deficit in notice-based governance.[438] When the scoring agreement is concluded with Zhima Credit Management Co., the rated person's contractual claims are channeled to the contracting scorer rather than to Taobao, even though the practical entitlement to use and operationalize the score is shared across the platform and its affiliated operators.[439] The resulting mismatch is visible in how informational responses are framed.[440] The platform's explanations for score decreases tend to foreground individualized breach narratives, such as late payment, and offer compliance tips, thereby steering the user toward self-correction while sidestepping potential scrutiny regarding whether the decline reflects model defect, data distortion, proxy discrimination, or policy choice.[441] This one-directional disclosure is therefore more accurately understood as an instrument of platform risk management and debt collection than one that meaningfully addresses the rated person's potential injury at the data processing and inference stages.[442]

The right to explanation is conceptually distinct from the right to be informed because it is not satisfied by one-way disclosure and evaluated by whether it closes a knowledge gap in a specific dispute.[443] Hilton's account of explanation treats it as a social information transfer process aimed at answering the explainee's questions and bridging knowledge asymmetry, rather than mere causal attribution.[444] In algorithmic decision-making, this view implies a

436. *See* Mahmood Bagheri & Chizu Nakajima, *Ex Ante and Ex Post Allocation of Risk of Illegality: Regulatory Sources of Contractual Failure and Issues of Corrective and Distributive Justice*, 13 EUR. J. L. & ECON. 5, 7 (2002) (discusses reconciling distributive and corrective justice in contract law, particularly in the allocation of risks associated with illegality).

437. *See id.* at 10–11 (discussing reconciling distributive and corrective justice in contract law, particularly in the allocation of risks associated with illegality).

438. *See* Xu Yong v. Zhima Credit Mgmt. Co., HANGZHOU INTERNET CT., (2018) Zhe 0192 Min Chu No. 302 (China).

439. *Id.*; *see also,* 黄刚诉浙江淘宝网络有限公司 [Huang Gang v. Zhejiang Taobao Network Co., Ltd.], (2019) Zhe 01 Min Zhong No. 7214, at 3–4, Hangzhou Interm. People's Ct. (China) (analyzing the Zhima Credit system); 王至宗诉浙江淘宝网络有限公司 [Wang Zhizong v. Zhejiang Taobao Network Co., Ltd.], (2018) Jing 0114 Min Chu No. 14773, Changping Dist. People's Ct. of Beijing (China) (analyzing the Zhima Credit system).

440. *Xu Yong*, Zhe 0192 Min Chu No. 302; *Huang Gang*, Zhe 01 Min Zhong No. 7214; *Wang Zhizong*, Jing 0114 Min Chu No. 14773.

441. *Xu Yong*, Zhe 0192 Min Chu No. 302; *Huang Gang*, Zhe 01 Min Zhong No. 7214; *Wang Zhizong*, Jing 0114 Min Chu No. 14773.

442. *See* Lu Bingbin (吕炳斌), *Geren Xinxi Baohu de "Tongyi" Kunjing jiqi Chulu* (个人信息保护的"同意"困境及其出路) [*The "Consent" Dilemma in Personal Information Protection and Its Solutions*], 38 FA SHANG YANJIU (法商研究) [STUD. L. & BUS.] 84, 87–90 (2021) (discussing how addressing the "consent" dilemma in personal information protection by distinguishing the intensity of consent based on the nature of information processing behaviors enhances explainability, as it provides clearer guidelines for individuals to understand and control how their data is used).

443. Wachter et al., *supra* note 88, at 4; *see also* Coeckelbergh, *supra* note 386, 2058–61 (discussing the moral problems in a lack of informed consent in explainable AI).

444. Denis J. Hilton, *Conversational Processes and Causal Explanation*, 107 PSYCH. BULL. 65, 65–67 (1990) (discussing how a conversational model of causal explanation enhances the explainability of algorithmic decisions by emphasizing the selection of explanations based on the focus of why questions, ensuring relevance

responsive duty.[445] Once a user is affected, the operator should analyze and communicate the logic chain, relevant factors, and how the result was produced in relation to the user's data, and this process should be capable of iterative refinement as users challenge and the system continues to learn.[446] This is why scholars describe the consent model as having structural and cognitive pathologies: while long, technical privacy policies create overload and routinized clicking, users lack the expertise to evaluate high-tech risks; consent becomes a ritual rather than a meaningful choice.[447] Therefore, the transition to an explanation right is a necessary correction if the law wishes to preserve the normative promise of informed consent in algorithmic environments.[448]

Moreover, EU law makes this distinction clear by combining static access rights with a narrower prohibition-style protection against certain automated decisions, indicating that information alone is insufficient.[449] Under Article 15(1)(h) GDPR, individuals subject to automated decision-making have the right of access.[450] Article 22 adds a further layer by addressing automated decisions that produce legal effects or similarly affect the data subject, which the WP29 interprets as including decisions that affect a person's financial situation, including eligibility for credit, job opportunities, or access to education.[451] Crucially, Article 22 can be read as a general prohibition rather than a right that must be individually invoked, meaning that individuals are automatically protected

and coherence in justifying outcomes; "Causal explanation is first and foremost a form of social interaction. One speaks of giving causal explanations … Someone explains something to someone").

445. Wachter et al., *supra* note 88, at 4; *see also* Wu, *supra* note 271, at 127–29 (outlining what a legal explainability requirement might use as criteria).

446. *See* Aimee Van Wynsberghe, *Sustainable AI: AI for Sustainability and the Sustainability of AI*, 1 AI & ETHICS 213, 213–15 (2021) (arguing that AI ethics must address explainability, responsibility, and governance across the lifecycle of AI systems rather than treating explanation as a purely technical afterthought).

447. *See* Jiang Su (江溯), *supra* note 301, at 84–85 ("Broad provisions do not yield satisfactory outcomes. For instance, when Google's search engine delivers intelligent ad placements, it complies with GDPR requirements by displaying "Why you might see this ad" to users. However, the disclosures remain rather generalised. Users cannot discern from these simplistic labels why specific advertisements appear before them, nor can they ascertain the precise weightings of the variables influencing automated decision-making"); *see* Haoran Li et al., *PrivaCI-Bench: Evaluating Privacy with Contextual Integrity and Legal Compliance*, PROCS. 63RD ANN. MEETING ASS'N FOR COMPUTATIONAL LINGUISTICS (Volume 1: Long Papers) 10544, 10544–46 (2025) (arguing that privacy evaluation should move beyond narrow personally identifiable information pattern matching and account for contextual information flows, including the sender, recipient, subject, transmitted attributes, purpose, consent, and transmission principle); *cf.* Wenbin Hu et al., *Context Reasoner: Incentivizing Reasoning Capability for Contextualized Privacy and Safety Compliance via Reinforcement Learning*, PROC. 2025 CONF. EMPIRICAL METHODS IN NAT. LANGUAGE PROCESSING 865, 865–68 (2025) (applying contextual integrity theory and reinforcement learning with rule-based rewards to improve LLM performance on privacy and safety compliance benchmarks grounded in the GDPR, EU AI Act, and HIPAA).

448. Wachter et al., *supra* note 88, at 4; *see also* Coeckelbergh, *supra* note 386, at 2058–61 (discussing the moral problems in a lack of informed consent in explainable AI).

449. *See, e.g.*, GDPR, art. 22 (addressing automated decisions that produce legal effects or similarly significantly affect the data subject); GDPR, art. 15 (securing the right of access for individuals); *see generally* Selbst & Poweles, *supra* note 56 (describing the function of the right to explanation under the GDPR).

450. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶ 7.

451. *See* Art. 29 Data Prot. Working Party, *Opinion 03/2013 on Purpose Limitation*, WP 203, at 45–47 (2013) (discussing the limits of information-based protection under the data protection framework).

regardless of whether they take action.[452] Furthermore, this design signals that the law does not assume that access requests and *ex ante* notice can carry the burden of protection.[453] The line of reasoning therefore supports recognizing the right to explanation as an independent, dynamic entitlement with rigid boundaries, rather than a thin subset of the static right to be informed.[454]

### *B. Is the right to explanation clearly recognized?*

In Europe, the idea of a right to algorithmic explanation is traced to the proposal that EU data protection rules should be read as requiring decision systems to be rendered contestable, where affected persons cannot realistically understand the operative logic, yet their informational position in the processing relationship must be protected.[455] Based on this account, the normative point is not technical transparency for its own sake but rather a duty on the designer or provider to communicate why, in terms intelligible to the data subject, the decision was reached and how it bears upon the subject's position.[456] Nevertheless, the GDPR does not expressly codify a freestanding right to explanation in Article 22, and this textual gap is the starting point for the academic dispute.[457] Article 22(3) grants procedural powers that presuppose intelligible reasons, since a data subject cannot express a meaningful view or contest a decision unless the controller discloses how the person's data contributed to the adverse result.[458] If controllers could comply through generic descriptions of system operation, then the safeguards in Article 22(3) would collapse into formality.[459] The core question is therefore functional coherence across the GDPR, not the presence or absence of a single formula in one provision.[460]

452. Florent Thouvenin et al., *Article 22 GDPR on Automated Individual Decision-Making: Prohibition or Data Subject Right?*, 2 EUR. DATA PROT. L. REV., 183, 184 (2022); *see generally* Luca, *supra* note 150 (discussing the ambiguity in Article 22 insofar as it defines a general prohibition or a right to be invoked).

453. Thouvenin et al., *supra* note 452, at 187; *see also* Selbst & Poweles, *supra* note 56, at 240–41 (explaining how *ex ante* and *ex post* distinctions ultimately convey a right to an explanation).

454. Wachter et al., *supra* note 88, at 4; Huanmin, *supra* note 53, at 1170; *see also* Selbst & Poweles, *supra* note 56, at 242 (concluding that the right to explanation exists within the GDPR as stated).

455. *See* Goodman & Flaxman, *supra* note 56, at 50–52, 55 (arguing that EU data protection law may be read as requiring algorithmic decisions to be contestable even where data subjects cannot meaningfully understand the underlying logic).

456. *See* Lin Huanmin (林洹民), *Geren Duikang Shangye Zidong Juece Suanfa de Siquan Sheji* (个人对抗商业自动决策算法的私权设计) [*The Design of Private Rights for Individuals Against Commercial Automated Decision-Making Algorithms*], 14 QINGHUA FAXUE TSINGHUA (清华学报) [U.L.J.] 125–127 (2020) (discussing how the design of private rights for individuals to counter commercial automated decision-making algorithms emphasizes the balance between protecting individual autonomy and ensuring the innovative value of algorithms).

457. GDPR, art. 22; *see generally* Luca, *supra* note 150 (discussing the lack of a distinct right in Article 22).

458. *See* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶ 55 ("[T]he main purpose of the data subject's right to obtain the information […] is to enable him or her effectively to exercise the rights conferred on him or her by Article 22(3) of that regulation"); GDPR, art. 22(3).

459. *See Dun*, ECLI:EU:C:2025:117, ¶¶ 59–60 (discussing what is ultimately necessary to operationalize the Article 22(3) right); *see also* Thouvenin et al., *supra* note 452, at 188 ("Article 22 GDPR must stipulate a data subject right. If automated decision-making was prohibited in general, the controller's duty to inform the data subject about the existence of automated decision-making would not make sense.").

460. Selbst & Poweles, *supra* note 56, at 242.

The weight carried by this recital depends on whether it is operationalized through enforceable provisions, as opposed to being treated as a freestanding entitlement, since recitals guide interpretation but do not—without more—create a complete cause of action.[461] On the one hand, a purposive and system-based reading treats Recital 71 as interpretive guidance whose practical force depends on it being operationalized through enforceable duties.[462] Based on this view, Articles 13 and 14 become decisive because they convert the recital's safeguards into concrete, source-sensitive disclosure obligations that require meaningful information about the logic involved as well as the significance and envisaged consequences for the data subject.[463] This approach is reinforced by reading Articles 13–15 together with Article 22, so that the quoted duties to disclose "*the existence of automated decision-making, [. . .], meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject*" are not treated as decorative transparency but as the informational substrate for challenge and correction.[464] The GDPR offers limited textual support for this framing in Recital 71, which ties the automated assessment of personal aspects, including the analysis or prediction of matters like economic situation, reliability, behavior, or location, to the need for safeguards, including the possibility of a right "*not to be subject to a decision*."[465] Recital 71 further supports this orientation by envisaging safeguards that include "*the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision*."[466] This helps to connect disclosure to contestation rather than to passive notice.[467]

Furthermore, a complementary strand frames Article 22 as more than an *ex post* remedy by pairing it with Article 25, so that data protection by design functions as an upstream design constraint across the system lifecycle, including bias-reducing choices in automated systems.[468] The same interpretive posture is sometimes supported by a rights of communication logic, since Article 10 of the *European Convention on Human Rights* provides that "*[e]veryone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers*"[469], and a broad reading of receiving information can be invoked to buttress the GDPR's demand for meaningful

461. *See* Kaminski, *supra* note 36, at 193–94 ("Recitals [have] a liminal legal status—they are not binding law, but they are often cited as authoritative interpretations where the GDPR is vague".).

462. *See* Selbst & Poweles, *supra* note 56, at 5 ("[I]t is sufficient to note that Recital 71 is not meaningless, and has a clear role in assisting interpretation and codetermining positive law.").

463. GDPR, arts. 13–14.

464. *Id*. art. 15; *Dun*, ECLI:EU:C:2025:117, ¶¶ 55–56.

465. GDPR, Recital 71.

466. *Id*.

467. Kaminski, *supra* note 36, at 204.

468. *See* Casey et al., *supra* note 87, at 180–81 (arguing that Article 22, when read together with Article 25 and the DPIA framework, operates not merely as an ex post remedy but as an upstream design constraint across the system lifecycle).

469. Convention for the Protection of Human Rights and Fundamental Freedoms art. 10, Nov. 4, 1950, 213 U.N.T.S. 221.

information in automated decision contexts (Article 10 ECHR).[470] Based on this account, explanation is instrumentally necessary because it enables the subsequent rights to express a view and to contest, and it also gives practical effect to the principles of lawfulness, fairness, and transparency in Article 5(1)(a) GDPR by ensuring that challenge is possible where the subject suspects substantive unfairness.[471]

On the other hand, a restrictive reading resists elevating this cluster of provisions into a broad explanation right;[472] instead, it treats the GDPR as supporting a narrow and largely post-decision entitlement.[473] Since the Recital 71 does not contain the stronger formulation occasionally attributed to it and that decision-specific explanations are more demanding than can be realistically supplied by system functionality in practice.[474] Based on this view, even when Articles 13–15 impose information duties, they are read as limited disclosures that fall short of compelling controllers to reveal detailed model structure, decision trees, criteria, or other elements that would approximate full transparency, due to which the claimed right remains thin in discrimination-sensitive settings.[475] A related line of disagreement concerns the legal character of Article 22(1), which lies between a *prima facie* prohibition on automated decision-making and a right that must be affirmatively exercised by the data subject;[476] the two readings allocate initiative and evidential burden very differently.[477] If Article 22(1) is a prohibition, controllers will rely on the exceptions within Article 22(2), including consent, to normalize the use of automation.[478] Meanwhile, if it is an opt-in right, the burden shifts to individuals to trigger non-automated alternatives, which may be unrealistic when services are standardized and dependency is high.[479]

In China, the legal basis for a right to algorithmic explanation is rooted in the PIPL.[480] Article 24(3) provides the following: "*Where a decision that has a major impact on an individual's rights and interests is made by means of automated decision-making [. . .] the individual shall have the right to request the personal information processor to make explanations [. . .] and refuse to accept*

470. Kaminski, *supra* note 36, at 207.

471. *See id*. ("[A]n individual has a right to explanation of an individual decision because that explanation is necessary for her to invoke the other rights—e.g., to contest a decision, to express her view. . ."); *See* Isak Mendoza & Lee A. Bygrave, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET L.: REGUL. AND ENF'T. 77, 87–88 (Tatiani Synodinou et al. eds., 2017) (explaining that the safeguards in Article 22(3) GDPR serve an instrumental function by enabling the rights to express a view and to contest automated decisions) (hereinafter Isak).

472. *See* Wachter et al., *supra* note 88 (arguing that the GDPR mandates only a limited right, not a right to the explanation of specific decisions).

473. *Id*.

474. *Id.* at 80 ("Recitals. . .are not themselves legally binding").

475. *Id*. at 15.

476. *See id.* at 19 (discussing the two interpretations of Article 22).

477. *See id.* at 20 (discussing the distinctions between the prohibition and right to object interpretations).

478. GDPR, art. 22; *see also id*. at 20 (discussing the prohibition view of Article 22(1)).

479. *See* Isak, *supra* note 471 ("The new exception for consent is likely to lower the de facto level of protection for individuals, particularly in light of the relative strength of most individuals vis-à-vis banks, insurance companies, online service providers and many other businesses.").

480. PIPL.

*that the personal information processor makes decisions solely by means of automated decision-making*."[481] Furthermore, Article 48 provides that "*[i]ndividuals shall have the right to request personal information processors to explain their personal information processing rules.*"[482] These provisions operate at the special and general levels, respectively, and provide a domestic statutory foundation for explanation claims in automated decision-making.[483] In addition, the Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services have developed an interpretability-oriented governance posture.[484] Article 12 provides that "*[a]n algorithm-recommended service provider shall be encouraged to comprehensively use content deduplication, intervention by shuffling, and other strategies, and optimize the transparency in and interpretability of search, sort, selection, push, and display rules [. . .].*"[485] Moreover, Article 17(3) provides that *"[i]f an algorithm-recommended service provider's application of algorithms exerts a significant impact on the rights and interests of users, it shall make an explanation and bear corresponding liability in accordance with the law*."[486] These provisions encourage providers to improve transparency and explainability while imposing a duty, where user rights and interests are materially affected, to provide explanations and assume responsibility.[487] These provisions also operate at the special and general levels, respectively, providing a statutory foundation for explanation claims in automated decision-making.[488]

Although Chinese scholarship continues to debate whether the right to algorithmic explanation has independent rights status, it is widely accepted as a supervisory procedure that supports the reasonable use of personal information, counters black-box decision-making, and advances transparency.[489] In practice, litigation in which individuals claim that their personal rights and interests have been affected by automated decision-making and seek judicial relief—because they have experienced confusion or harm regarding the decision's manner and logic—has already begun to proliferate.[490] In this context, the right to algorithmic explanation functions as a communicative mechanism between the subject

481. *Id.* art. 24(3).

482. *Id.* art. 48.

483. *Id.* arts 24(3), 48.

484. Regulations on the Management of Internet Information Service Algorithm Recommendation (互联网信息服务算法推荐管理规定) (promulgated Dec. 31, 2021, effective Mar. 1, 2022) [hereinafter PIPL Recommendations].

485. *Id.* art. 12.

486. *Id.* art. 17(3).

487. *Id*. arts 12, 17(3).

488. *Id.*

489. *See generally* Huanmin, *supra* note 53 (discussing the right to explanation in Chinese law and the implications for addressing the opacity of algorithmic decision-making in China).

490. *See* Chen v. Hangzhou Alimama Software Service Co., Hangzhou Railway Transp. Ct., (2017) Zhe 8601 Min Chu No. 3306 (China) (involving a challenge to platform-imposed sanctions where the claimant alleged confusion and harm arising from opaque automated assessment and the inability to meaningfully understand or contest the decision-making logic); *see also* Zhao v. Zhejiang Taobao Network Co., Hangzhou Internet Ct., (2020) Zhe 0192 Min Chu No. 2295 (China); Ling v. A Short Video Co., Beijing Internet Ct., (2019) Jing 0491 Min Chu No. 6694 (China).

and the processor clarifying how the algorithm operates.[491] It plays an irreplaceable institutional role in rebuilding trust and achieving a normatively defensible balance of rights and powers.[492] Only by granting the right to obtain detailed explanations after the event to users can adjudication reflect the rated person's information property interests adequately, thereby incentivizing platforms to assume transparency and accountability responsibilities while pursuing commercial gain.[493]

### *C. The U.S. Adverse-Action Model of Credit Explanation*

The European and Chinese materials discussed above frame an explanation through data protection, automated decision-making, and the relationship between the data subject and the processor.[494] The United States instead builds its concrete explanation duties around the moment of adverse credit treatment.[495] On the one hand, this design is narrower than the GDPR or the PIPL, as it does not create a general right to understand automated decision-making throughout all contexts.[496] On the other hand, it is also more operational in one respect.[497] Credit denial, unfavorable pricing, account termination, credit line reduction, or refusal to increase a credit limit constitutes a legally regulated reason-giving event.[498] In the United States, credit scoring evolved from standardized scorecards and bureau-based credit reports to a more extensive risk infrastructure that includes credit bureaus, lenders, securitization markets, fintech firms, alternative data providers, and third-party model vendors.[499] The ECOA provides that "*[a] statement of reasons meets the requirements of this section only if it contains the specific reasons for the adverse action taken*."[500] Regulation B provides more

491. *See* Huanmin, *supra* note 53, at 1190 ("Accordingly, the government should be obligated to explain algorithmic decisions because an individual is not capable of reading and analysing algorithms and cannot understand algorithmic decisions without the help of the person using the algorithm.").

492. *See id.* at 1191 (discussing how the right to explanation can help balance individual, public, and commercial interests).

493. *See* Grochowski et al., *supra* note 39, at 58, 61 (discussing transparency and explainability).

494. *See supra* notes 463–466, 480–482 and accompanying text (discussing the GDPR and PIPL's provisions).

495. 12 C.F.R. § 1002.9(a)(2).

496. *See* Kaminski, *supra* note 36, at 191–92 ("The literature in the United States has been largely speculative, operating in a policy vacuum. This is resolutely not, however, the case in the European Union."); *see also* Emily Gross, *Behind the Screen: A Comparison of Data Protection Regulations in the EU and US*, INT'L & COMPAR. L. REV. (Dec. 4, 2023) https://international-and-comparative-law-review.law.miami.edu/behind-the-screen-a-comparison-of-data-protection-regulations-in-the-eu-and-us/ [https://perma.cc/B3XZ-ZP3X] (discussing the narrowness of the US approach compared to the EU).

497. 12 C.F.R. § 1002.9(b)(2).

498. *See* Equal Credit Opportunity Act, 15 U.S.C. § 1691(d) (discussing what constitutes an adverse action); *see also* Citron & Pasquale, supra note 24, at 15–16 (discussing credit denial reasoning).

499. *See* Citron & Pasquale, *supra* note 24, at 11–16 (describing the rise of credit scoring and the social consequences of scoring systems); *see* Scorecards as Devices for Consumer Credit, *supra* note 63, at 288–93 (tracing the development of FICO-style credit scorecards); Andreas Fuster et al., *The Role of Technology in Mortgage Lending*, 32 REV. FIN. STUD. 1854, 1854–61 (2019) (examining technology-driven changes in mortgage lending); *see* Berg et al., *supra* note 9, at 2885–86 (discussing digital footprints and how they complement information from credit bureaus).

500. Equal Credit Opportunity Act, 15 U.S.C. § 1691(d)(3).

detailed content for the duty.[501] Section 1002.9(a)(2)(i) requires notice of adverse action to contain either "*a statement of specific reasons for the action taken*" or notice of the applicant's right to obtain such reasons.[502] Section 1002.9(b)(2) further provides that "*[t]he statement of reasons for adverse action [. . .] must be specific and indicate the principal reason(s) for the adverse action.*"[503] The same provision rejects generic explanations by stating that "*[s]tatements that the adverse action was based on the creditor's internal standards or policies or that the applicant [. . .] failed to achieve a qualifying score on the creditor's credit scoring system are insufficient.*"[504] The official interpretation further requires that the disclosed reasons correspond to and accurately describe the factors the creditor actually considered or scored.[505] In credit-scoring cases, the creditor may not omit any principal reason for adverse action and must identify the actual basis for denial, even where the connection between that factor and creditworthiness would not be obvious to the applicant.[506]

Earlier U.S. regulatory guidance illustrates the relevance of this model to algorithmic credit scoring.[507] CFPB Circular 2022-03 stated that ECOA and Regulation B required creditors to make specific reasons for adverse action available, even when complex algorithms made the reasons for denial difficult to identify.[508] The Bureau's position went beyond a general expectation that algorithmic creditors should attempt to explain their decisions.[509] On the Bureau's reading of ECOA and Regulation B, creditors could not rely on complex algorithms that prevented disclosure of specific and accurate reasons for adverse action.[510] Model complexity and limited institutional understanding left the creditor's responsibility intact, including the need to validate any post hoc approximation for accuracy.[511]

CFPB Circular 2023-03 addressed a related concern: generic notices that obscured the actual grounds for adverse action.[512] The Circular examined whether sample-form checklists could suffice when the listed reasons failed to

501. 12 C.F.R. § 1002.9(b).

502. *Id*. § 1002.9 (a)(2)(i).

503. *Id.* § 1002.9(b)(2).

504. *Id.*

505. *Id.* § 1002.9(b)(2)-(4).

506. *Id.*

507. *See* CFPB 2022, *supra* note 51; *see also CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence*, CONSUMER FIN. PROT. BUREAU (Sep. 19, 2023) https://www.consumerfinance.gov/archive/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/ [https://perma.cc/URG2-23G9 ] (discussing regulatory triggers for AI driven credit scoring); Circulars 2022-03 and 2023-03 were withdrawn on May 12, 2025 and are discussed here as historical regulatory interpretations. *See* Interpretive Rules, Policy Statements, and Advisory Opinions; Withdrawal, 90 Fed. Reg. 20084 (May 12, 2025).

508. *See* CFPB 2022, *supra* note 51. ("ECOA and Regulation B require creditors to provide statements of specific reasons to applicants against whom adverse action is taken.").

509. *Id.*

510. *See id.* ("ECOA and Regulation B do not permit creditors to use complex algorithms when doing so means they cannot provide the specific and accurate reasons for adverse actions.").

511. *Id.*

512. *See* CFPB 2023, *supra* note 51. ("Nor, as a general matter, may creditors rely on overly broad or vague reasons to the extent that they obscure the specific and accurate reasons relied upon.").

identify the actual basis for the decision.[513] The CFPB's answer was no.[514] Selected reasons had to identify the principal grounds accurately and specifically, without broad or vague categories obscuring the decision's actual basis.[515] The Circular emphasized the risk posed by models drawing on data outside conventional credit applications or files.[516] For example, the Circular cautioned that a denial based on an applicant's profession could require more specific explanation than a reference to insufficient projected income.[517] The Circular also treated a generic reference to purchasing history as potentially inadequate when a creditor reduced an existing credit line using behavioral data about the establishments a consumer patronized or the goods purchased.[518] On that interpretation, a legally adequate explanation had to identify the principal reasons for adverse action without obscuring those reasons through a vague consumer-facing category.[519] Although withdrawn by the CFPB on May 12, 2025, both Circulars illustrate the Bureau's earlier efforts to hold creditors responsible for accurate and specific explanations of algorithmic credit decisions.

FCRA complements this model by tying adverse-action notice to consumer-report information and credit-score factors.[520] Section 1681m(a) provides that if a person takes adverse action "based in whole or in part on any information contained in a consumer report," that person must provide notice of the adverse action and identify the consumer reporting agency that furnished the report.[521] Where a credit score is used, Section 1681m(a)(2) requires disclosure of the numerical credit score and related information required by Section 1681g(f).[522] Section 1681g(f)(1)(C) requires disclosure of "all of the key factors that adversely affected the credit score of the consumer in the model used," generally limited to four factors, or five where inquiries are one of the key factors.[523] Section 1681g(f)(2)(B) defines key factors as *"all relevant elements or reasons adversely affecting the credit score of the particular individual, listed in the order of their importance based on their effect on the credit score.*"[524] FCRA also

513. *See id.* ("When using artificial intelligence or complex credit models, may creditors rely on the checklist of reasons provided in CFPB sample forms for adverse action notices even when those sample reasons do not accurately or specifically identify the reasons for the adverse action?").

514. *Id.*

515. *See id.* ("No, creditors may not rely on the checklist of reasons provided in the sample forms (currently codified in Regulation B) to satisfy their obligations under ECOA if those reasons do not specifically and accurately indicate the principal reason(s) for the adverse action. Nor, as a general matter, may creditors rely on overly broad or vague reasons to the extent that they obscure the specific and accurate reasons relied upon.").

516. *See id.* ("Building on that previous guidance, this Circular focuses on the accuracy and specificity requirements of those notices, even when such models, driven by data gathered outside of traditional credit reports or applications, are utilized.").

517. *See id.* ("For instance, if a complex algorithm results in a denial of a credit application due to an applicant's chosen profession, a statement that the applicant had 'insufficient projected income' or 'income insufficient for amount of credit requested' would likely fail to meet the creditor's legal obligations.")

518. *Id.*; *see also* Complaint at 34–35, FTC v. CompuCredit Corp., No. 1:08-cv-1976-BBM-RGV (N.D. Ga. June 10, 2008) (alleging use of behavioral data to reduce credit lines).

519. *Id.*; CFPB 2023, *supra* note 51.

520. FCRA, 15 U.S.C. § 1681m(a).

521. *Id.* § 1681m(a)(1)–(3).

522. *Id.* § 1681m(a)(2)(A)–(B); 15 U.S.C. § 1681g(f)(1)(B)–(E).

523. 15 U.S.C. § 1681g(f)(1)(C).

524. 15 U.S.C. § 1681g(f)(2)(B).

gives the consumer access and dispute rights, including the right to obtain file information and to dispute incomplete or inaccurate information.[525] The resulting structure is a correction-oriented notice regime linking an adverse decision to the report, the score, the key adverse factors, and the consumer's right to obtain and dispute the underlying file.[526] Besides, the case law confirms the procedural function of an adverse action notice.[527] The *Fischl v. General Motors Acceptance Corp.* described the ECOA notice requirement as one of the most significant amendments to ECOA and stressed its consumer-protection and educational purposes.[528] In *Treadway v. Gateway Chevrolet Oldsmobile Inc.,* the Seventh Circuit stated that the notice requirement discourages discrimination and educates consumers about deficiencies in their credit status.[529]

Nevertheless, the U.S. model exposes its own institutional gaps.[530] First, adverse-action notices may be sufficiently specific to satisfy Regulation B while still too thin to explain machine-learning interaction effects.[531] Regulation B provides that a creditor "need not describe how or why a factor adversely affected an applicant."[532] A notice may therefore identify length of residence, age of automobile, or another scored factor without explaining weight, interaction, proxy effect, or causal role.[533] Second, FCRA key-factor disclosure is limited because it provides consumers with adverse score factors without requiring disclosure of the scoring formula or how credit history is converted into a score.[534] Third, coverage depends on statutory categories such as consumer report and consumer reporting agency.[535] Alternative-data providers, device-level signals, household-level inferences, marketing lists, and internal lender analytics may fall outside FCRA or create uncertainty at the boundary.[536] Fourth, trade-secret

525. *Id.* §§ 1681g(c), 1681j.

526. *See* Hurley & Adebayo, *supra* note 16, at 170–75 (explaining that FCRA seeks to give consumers access to credit information, correction mechanisms, and information about how credit data are used by third parties); *see* Citron & Pasquale, *supra* note 24, at 21–25 (discussing FCRA's credit-score disclosures, key-factor requirements, and the continued opacity of how credit histories are converted into scores).

527. Treadway v. Gateway Chevrolet Oldsmobile Inc., 362 F.3d 971, 977–78 (7th Cir. 2004); Fischl v. Gen. Motors Acceptance Corp., 708 F.2d 143, 146 (5th Cir. 1983);

528. *See Fischl*, 708 F.2d at 146 ("Perhaps the most significant of the 1976 amendments to the ECOA, these provisions were designed to fulfill the twin goals of consumer protection and education.").

529. *See Treadway*, 362 F.3d at 977 ("In sum, the notice requirement serves two purposes: it discourages discrimination and it educates consumers as to the deficiencies in their credit status.").

530. 12 C.F.R. pt. 1002, Supp. I, § 1002.9; *see generally* Sahiba Chopra, *Current Regulatory Challenges in Consumer Credit Scoring Using Alternative Data-Driven Methodologies*, 23 VANDERBILT J. OF ENT. & TECH. L. 625 (2021) (discussing some of these institutional regulatory gaps).

531. 12 C.F.R. pt. 1002, Supp. I, § 1002.9; CFPB 2022, *supra* note 51.

532. 12 C.F.R. pt. 1002, Supp. I, § 1002.9(b)(2)–(3).

533. *Id.* at para. 9(b)(2)(3).

534. 15 U.S.C. §§ 1681g(f)(1)(C), 1681m(a)(2) (2026); *see* Citron & Pasquale, *supra* note 24, at 17 ("The law does not require credit scorers to tell individuals how much any given factor mattered to a particular score.")

535. 15 U.S.C. § 1681a(d)–(e); *see* Hurley & Adebayo, *supra* note 16, at 170–78 (explaining that FCRA seeks to give consumers access to credit information, correction mechanisms, and information about how credit data are used by third parties); *see also* McCready v. eBay, Inc., 453 F.3d 882, 891–92 (7th Cir. 2006) (holding that information associated with an anonymous username did not qualify as a consumer report about an identifiable consumer); *see* Fuges v. Sw. Fin. Servs., Ltd., 707 F.3d 241, 253–55 (3d Cir. 2012) (addressing the limits of FCRA coverage where a report concerned property-related information).

536. *See* Hurley & Adebayo, *supra* note 16, at 180–84; *see also McCready*, 453 F.3d at 889 (holding that information associated with an anonymous username did not qualify as a consumer report about an identifiable

and vendor-control structures can preserve the confidentiality of scoring models, while leaving the affected consumer with only factor labels.[537] These limits can each prevent a consumer from converting notice into contestation.[538]

Another important fault is that access, notice, and explanation frequently remain detached from the moment when the borrower actually suffers adverse credit treatment.[539] GDPR-style access may remain too general when no concrete credit decision is identified.[540] PIPL-style explanation may also lose practical force when platform responses remain one-way, generic, or non-reviewable.[541] The U.S. adverse-action model usefully shifts attention to the decision point, but its own weakness is that factor labels can satisfy notice duties while leaving proxy effects, alternative data, or vendor-controlled scoring practically untested.[542] For that reason, this article suggests a reform path.[543] A more useful notice would identify the principal reason relied upon, specify whether the reason came from a credit report, alternative data source, or internal behavioral signal, indicate whether automated scoring was decisive, and state the route for correction or human review.[544] For example, if a credit line is reduced because spending data from certain merchant categories is treated as a negative signal, the notice should not stop at the purchasing history alone.[545] It should identify the spending signal, its source, and the route for correction or review.[546] This

consumer); *Fuges*, 707 F.3d at 249–50 (addressing the limits of FCRA coverage where a report concerned property-related information).

537. *See* Fair Isaac Corp. v. Experian Info. Sols., Inc., 645 F. Supp. 2d 734, 739–43 (D. Minn. 2009), aff'd, 650 F.3d 1139 (8th Cir. 2011) (describing disputes over credit-scoring systems developed by FICO and credit bureaus); *see* Citron & Pasquale, *supra* note 24, at 28–30 (discussing FCRA's credit-score disclosures, key-factor requirements, and the continued opacity of how credit histories are converted into scores).

538. *See* Reynolds v. Hartford Fin. Servs. Grp., Inc., 416 F.3d 1097, 1109–10 (9th Cir. 2005) ("Providing notice therefore serves the statutory purpose of allowing the consumer to correct errors in credit reports."); *see also* Citron & Pasquale, *supra* note 24, at 17–19 ("those four factors do little to explain credit scores.").

539. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 53–56 (explaining that access and explanation serve to enable affected persons to understand, challenge, and seek correction of automated decisions); *see generally* Cameron Huddleston, *5 Ways a Bad Credit Report Can Hurt You*, KIPLINGER (Dec. 8, 2023) (describing some of the risks of adverse credit treatment).

540. Case C-634/21, OQ v. Land Hessen, EU:C:2023:957, ¶¶ 61–63 (identifying a potential "lacuna in legal protection" where a credit-scoring entity generates the score but a third party makes the ultimate decision); *see* Hurley & Adebayo, *supra* note 16, at 170–78 (explaining that FCRA seeks to give consumers access to credit information, correction mechanisms, and information about how credit data are used by third parties).

541. Huanmin, *supra* note 53, at 1188–89 (distinguishing decision-specific explanations from the more general algorithmic information available in the individual–enterprise context); PIPL, art. 24 (2021).

542. *See* 12 C.F.R. pt. 1002, Supp. I, § 1002.9, para. 9(b)(2)–(3); Hurley & Adebayo, *supra* note 16, at 170–87.

543. Hurley & Adebayo, *supra* note 16, at 149.

544. *See* CFPB 2022, *supra* note 51 (explaining that reasons must accurately identify the factors actually considered or scored and that notice enables consumers to rectify mistakes); *see also* 12 C.F.R. § 1002.9(b)(2) (requiring specific principal reasons for adverse action).

545. *See* CFPB 2023, *supra* note 51 ("For example, if a creditor decides to lower the limit on, or close altogether, a consumer's credit line based on behavioral data, such as the type of establishment at which a consumer shops or the type of goods purchased, it would likely be insufficient for the creditor to simply state "purchasing history" or "disfavored business patronage" as the principal reason for adverse action.").

546. *See id.* ("Instead, the creditor would likely need to disclose more specific details about the consumer's purchasing history or patronage that led to the reduction or closure, such as the type of establishment, the location of the business, the type of goods purchased, or other relevant considerations, as appropriate.").

would not require full access to algorithmic logic, but would connect reason-giving, data correction, and review where contestability matters most.[547]

## VI. LEGAL CONSTRUCTION OF THE RIGHT TO EXPLANATION

The right to an algorithmic explanation is often understood as a user's right to know why the system reached a particular conclusion, potentially creating the misleading impression that the AI tool itself must justify its output.[548] This intuition is legally and technically unstable.[549] The more defensible starting point is to treat explanation as a contestability device that enables the affected person to test whether the negative score is lawful, fair, and factually sound, rather than treating it as a fabricated internal psychology of the model.[550] Once it is treated as a contestability mechanism, the key doctrinal task is to construct explanation in such a manner that it can engage in credit-scoring disputes while remaining bounded and administrable.[551] For the duty bearer, the primary obligation should rest on the scoring operator that determines purposes and means, because it controls the decision pipeline and can implement remediation, whereas the developer should bear an assistance duty where technical knowledge is necessary to translate system operations into an intelligible external account.[552] Regarding the standard and mode, the explanation must be given in such a form that the subject can understand and decide whether and how to seek remedies while maintaining appropriate (rather than total) transparency to avoid converting the explanation into a disclosure of technical secrets.[553] Accordingly, the essence of

547. *See* Citron & Pasquale, *supra* note 24, at 28 ("Aggrieved consumers could be guaranteed reasonable notice if scoring systems included audit trails recording the correlations and inferences made algorithmically in the prediction process."); *see also* Hurley & Adebayo, *supra* note 16, at 170–78 (explaining that FCRA seeks to give consumers access to credit information, correction mechanisms, and information about how credit data are used by third parties).

548. *See* Wachter et al., *supra* note 356, at 845 ("In the existing literature, 'explanation' typically refers to an attempt to convey the internal state or logic of an algorithm that leads to a decision."); Jason M. Loring, *The Explainability Illusion: Why AI Transparency Requirements Miss the Point*, JONES WALKER (Aug. 11, 2025) https://www.joneswalker.com/en/insights/blogs/ai-law-blog/the-explainability-illusion-why-ai-transparency-requirements-miss-the-point.html?id=102kzle [https://perma.cc/H66E-ATBA].

549. *See* Wachter et al., *supra* note 356, at 863 ("Attempts to implement a right to explanation that opens the 'black box' to provide insight into the internal decision-making process of algorithms face four major legal and technical barriers.").

550. *See* Wachter et al., *supra* note 356, at 863 ("This approach provides data subjects with meaningful explanations to understand a given decision, grounds to contest it, and advice on how the data subject can change his or her behaviour or situation to possibly receive a desired decision in the future.").

551. *See* Kars Alfrink et al., *Contestable AI by Design: Towards a Framework*, 33 MINDS & MACHINES 613, 614-15 (2023) (conceptualizing contestability as a procedural relationship requiring meaningful opportunities to challenge automated decisions); *see also* Xinbao Zhang (张新宝), *supra* note 19, at 72 (arguing that the primary duty to provide explanation rests with the automated decision-making operator, while developers bear an auxiliary obligation to assist where technical expertise is required).

552. *See* Linghan Zhang (张凌寒), *supra* note 26, at 72 (arguing that the primary duty to provide explanation rests with the automated decision-making operator, while developers bear an auxiliary obligation to assist where technical expertise is required).

553. *See* Zhang Endian (张恩典), *Dà Shù Jù Shí Dài De Suàn fǎ Fiě Shì Quán: Bèi Jīng, Luó Jí Yǔ Gòu Zào* (大数据时代的算法解释权：背景、逻辑与构造) [*Background, Logic and Structure of the Right to Explanation of Algorithmic Decision-Making in the Age of Big Data*], 34 FǍ XUÉ LÙN TÁN (法学论坛) [LEGAL F.] 152, 157–58 (2019) (discussing how the right to explanation for algorithmic decisions, aims to enhance

explanation is a dynamic disclosure duty concerning system structure, operational mechanisms, and bias risk.[554]

### *A. The right holder and duty bearer*

If a controller chooses to govern credit access through automated scoring, it cannot treat the model as an excuse or a shield.[555] That is because a risky instrument entails a supervisory role.[556] This intuition aligns with a legal structure, where private autonomy and responsibility track rational capacity.[557] For this reason, law separates the capacity to perform legal acts from the capacity to bear tort responsibility and uses a guardian-style supervision duty as a necessary supplement when the primary actor cannot reliably govern itself.[558] The same structure can be transposed to algorithmic scoring, not by personifying the AI tool, but by assigning supervision and reason-giving duties to the human and corporate actors who deploy it.[559] The GDPR and the PIPL frame personal data protection around the natural person as the protected subject.[560] The GDPR aims to protect natural persons, defining personal data as any data that relate to an identified or identifiable natural person.[561] The PIPL takes the same direction in Article 2 and subsequently triggers the explanation-relevant safeguard when an automated decision significantly impacts an individual.[562] The right holder should therefore be the natural person who is the counterparty of the automated scoring outcome, not a firm or an abstract category of users, because the function of explanation is ensuring contestability in a concrete adverse impact.[563]

The more difficult question concerns the duty bearer.[564] While the guardian analogy helps justify this allocation, it also sets limits.[565] A guardian's duty encompasses protection and control, and Article 34 of the Civil Code describes

transparency and accountability by granting data subjects the right to understand the logic behind algorithmic decisions).

554. *Id.*

555. *See* Fowler V. Harper & Posey M. Kime, *The Duty to Control the Conduct of Another*, 43 YALE L.J. 886–89 (1934) [hereinafter Harper] (explaining Judge Cardozo's theory that the conduct of one that brings them into a human relationship with another owes the person a duty of care, even if it is through the use of the defendant's chattel).

556. *Id.* at 892 (highlighting the danger of an activity is a fact of importance in determining the tort liability of an operator).

557. *See e.g.*, *id.* at 892–95 (tort liability extends to operators of vehicle and guardians of children because vehicles and children lack rational capacity to create harm).

558. *See id*. at 886–94 (arguing that tort responsibility turns on relational control capacity rather than formal status, and that affirmative supervisory duties arise where the primary actor lacks the ability to self-govern).

559. *See generally*, Miriam Buiten et al., *The Law and Economics of AI Lability*, 48 COMP. L. & SEC. REV. 1, 8 (2023) (explaining the different ways regulations can supervise AI).

560. GDPR, art. 1; PIPL, art. 2.

561. GDPR, art. 1.

562. PIPL, art. 2.

563. *See* Emre Bayamlıoğlu, *The Right to Contest Automated Decisions Under the General Data Protection Regulation: Beyond the So-Called "Right to Explanation"*, 16 REG. & GOV'T 1058, 1066–68 (2022) (discussing how the right to contest automated decisions under the GDPR enhances algorithmic explainability by mandating transparency measures that facilitate meaningful contestation) [hereinafter Right to Contest].

564. *See* Buiten et al., *supra* note 559, at 19 (discussing the difficulty in correctly assigning duty).

565. *See* Chinese Civil Code, art. 34.

this baseline as follows: "The duties of a guardian are to represent the ward to perform civil juristic acts and to protect the personal, proprietary, and other lawful rights and interests of the ward."[566] This structure highlights two notable points.[567] First, supervision is not passive, because a guardian must manage foreseeable risk rather than apologizing after harm, and second, supervision is not absolute confinement, because excessive control can undermine legitimate autonomy.[568] For algorithmic scoring, these points imply a bounded but real supervisory package.[569] The primary duty bearer must be the actor with operational control, because it can implement safeguards, adjust variables, and build a human review path.[570] Meanwhile, the developer should bear an assistance duty where technical knowledge is needed to translate procedures and principles into an intelligible external account.[571] Where control and use are separated, a joint or linked duty structure is required to ensure that explanation does not fail when outsourcing is used.[572] The guardian model also cannot be imported mechanically as carrying strict liability for all AI-related harms, where the right to explanation must be kept distinct from tort allocation.[573] However, AI does not simply resemble a person who lacks capacity.[574] While it lacks general human understanding, in some domains, it performs at a level that exceeds many fully capable adults;[575] therefore, it has already been deployed widely in robo-advisories, autonomous driving, and digital assistants, and large models are increasingly displaying human-like task execution.[576] Treating AI as equivalent to a person with limited to no capacity ignores this domain-specific competence and distorts incentives by pushing all losses onto the controller, regardless of realistic

566. *Id.*

567. *See* Harper, *supra* note 555, at 894 (expounding on the affirmative duty of the parent to control the child's habits that may cause harm on others); *see* Chinese Civil Code, art. 35 (providing the duties and interests of the guardian must relate to the minor's interests and true will based on their age and intelligence).

568. *Id.*

569. Right to Contest, *supra* note 563, at 1074 (calling for a comprehensive compliance transparency scheme that will help define and treat the risks to the rights of data subjects); *see* Oreoluwa Onabowale, *The Rise of AI and Robo-Advisors: Redefining Financial Strategies in the Digital Age*, 6 INT'L J. RSCH. PUB. & REVS. 4832, 4839 (highlighting that there's existing algorithmic biases in credit scoring that require continuous monitoring).

570. Right to Contest, *supra* note 563, at 1072 (explaining the need role tools that scrutinize automated decisions can have in discovering features that are relevant to algorithmic decisions, identify values that needs to be altered to change outcomes); *see also* Buiten et al., *supra* note 559, at 11–13 (discussing different AI liability theories based on capacity or lack of control).

571. *See* Florence Millerand & Karen S. Baker, *Who Are the Users? Who Are the Developers? Webs of Users and Developers in the Development Process of a Technical Standard*, 20 INFO. SYS. J. 137, 150–58 (2010) (discussing how the development process of technical standards, such as the Ecological Metadata Language, involves dynamic webs of users and developers whose evolving roles and interactions are crucial for ensuring the practicality and adaptability of standards).

572. *See* Buiten et al., *supra* note 559, at 20 (arguing regulation should hold both producers and operators liable as care by each party is essential to avoid algorithmic failure).

573. *Id.* at 3 (explaining situations in which strict liability regimes are justified).

574. *See* Onabowale, *supra* note 569, at 4835 (examining the important role AI has in analyzing data in financial automation).

575. *Id.*

576. *Id.*

precautions.[577] This article, therefore, rejects a general move toward no-fault liability or presumed fault for AI use.[578] Article 86 of the EU AI Act places the duty on the deployer as follows:

> Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system. . ., and which produces legal effects or similarly significantly affects that person. . .shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.[579]

Although this looks like a way out of the GDPR style impasse, it risks flattening the AI lifecycle into a single point of accountability.[580] This is because deployers often lack access to the provider-side information required to explain how outputs were generated.[581] Without an explicitly integrated duty that pulls in the provider where necessary, Article 86 risks becoming a formal remedy that shifts the burden to the least informed actor and encourages misplaced confidence that explanation alone is sufficient.[582]

Another text-level answer is that the PIPL assigns the obligation to the personal information processor,[583] and Article 24 paragraph 3 is triggered when a decision made through automated decision-making has a significant impact on individual rights and interests,[584] while Article 48 provides the correction pathway when an obligation is breached.[585] However, the institutional problem is that automated credit scoring is usually produced by a chain rather than a single actor, and the contract counterparty is often not the actor who actually controls the model, variables, or data flows.[586] In practice, large platforms concentrate control through data advantage and model ownership, while a bank or a public body may only call the service.[587] This makes a purely formal duty allocation

577. *Contra* Emiliano Marchisio, *In Support Of "No-Fault" Civil Liability Rules For Artificial Intelligence*, 1 SN SOC. SCI., Jan 2021 (examining that current liability structure confounds the fault of developers and users).

578. *Id.*

579. EU AI Act, art. 86.

580. *See* Julie Owono, *AI Needs Accountability. We Can't Rely on Companies and Governments Alone*, JUST SEC. (Mar. 27, 2026), https://www.justsecurity.org/133322/ai-accountability-companies-governments/ [https://perma.cc/RDK7-DNRY] (arguing that proper accountability for AI "should not consist of a single entity, but an ecosystem of layered accountability").

581. *See* EU AI Act, art. 13 (providing a requirement to maintain a degree of transparency of information to deployers)

582. *See* Arnoud Engelfriet, *An Uninterpretable Right: Legal and Practical Limits of the Right to an Explanation*, VRIJE UNIVERSITEIT AMSTERDAM, June 2025 (discussing the limits of the European Union's AI Act because explanations can create an illusion of control).

583. PIPL, art. 24.

584. *Id.*

585. *Id.* art. 48.

586. *See* Nizan G. Packin & Yafit Lev-Aretz, *Decentralized Credit Scoring: Black Box 3.0*, 61 AM. BUS. L. J. 91, 95 (2024) (discussing how decentralized credit scoring systems, by integrating off-chain data with blockchain-based smart contracts, introduce unique challenges to algorithmic explainability due to their reliance on complex smart contract logic and external oracle data; "Robo-advisors . . . have evolved to incorporate sophisticated capabilities such as machine learning and natural language processing, enabling increasingly customized and adaptive recommendations.").

587. *Id.*

under inclusive terms of service incapable of delivering a real explanation.[588] A further complication is corporate structuring.[589] A group may place the user contract in a thin subsidiary while the core platform retains the scoring infrastructure and downstream reuse of the score, which turns contractual privity into a liability firewall.[590] The correct principle is therefore correspondence between power and responsibility.[591] Where the scoring operator is only a front end, the actor that actually determines purposes and means must bear a primary explanation duty, and the contract counterparty must at least bear a duty to coordinate and deliver an explanation that meets the statutory standard, rather than pointing users to internal complexity.[592] In regulated credit markets, duty allocation should turn on functional control over the adverse decision.[593] The lender or deployer that communicates or implements denial, repricing, credit-line reduction, or account termination should ordinarily bear the consumer-facing explanation duty.[594] Such an approach is consistent with the U.S. adverse-action model, which assigns the reason-giving duty to the creditor taking adverse action and does not permit model complexity to excuse inadequate explanation.[595]

In addition, the scoring provider or model developer should bear an assistance duty where the deployer cannot produce a meaningful explanation without technical information.[596] Without such an assistance duty, Article 86 of the EU AI Act risks placing the outward explanation burden on the deployer while leaving the information needed for that explanation with the provider;[597] PIPL-style processor obligations face a similar problem where the contracting platform is not the actor that actually controls the scoring infrastructure.[598] In this sense, the right to explanation should not be constructed as a demand that a consumer personally inspect the whole model.[599] Where a bank relies on an internal ratings-based system, automated creditworthiness model, or vendor scoring tool, and adverse-action reason mappings should be preserved so that regulators, courts,

588. *Id.* at 92.

589. *See* TIM MATHER ET AL., CLOUD SECURITY AND PRIVACY: AN ENTERPRISE PERSPECTIVE ON RISKS AND COMPLIANCE (2009) (describing multi-layer enterprise architectures and distributed operational control).

590. *Id.*

591. *See* Harper, *supra* note 555, at 886–94; *cf.* Guoying Lu et al., *AI as Skills Governance: Reconfiguring Authority and Ethics in China's TVET Reform*, 5 ASIAN REV. POL. ECON., art. 12, at 13–14, 20 (2026) (arguing that accountability in platform-based vocational education should align with actual decision-making power, given the separation between educators' formal responsibility and platform control over evaluation criteria).

592. *See e.g.*, EU AI Act, art. 50 (requiring separate responsibilities for providers to ensure AI has recognizable features and requiring deployers to inform the use of data).

593. *See* Harper, *supra* note 555, at 886–894.

594. *See E*qual Credit Opportunity Act, 15 U.S.C. § 1691(d)(2)–(3); 12 C.F.R. §§ 1002.9(a)(2), (b)(2); CFPB 2022, *supra* note 51 (requiring creditors to explain adverse action and information pertaining to the adverse action).

595. *Id.*

596. *See, e.g.*, EU AI Act, art. 53 (requiring providers to maintain and comply with providing necessary information).

597. *Id.* art. 86.

598. *See* PIPL, arts. 21, 24, 73 (1)–(2) (2021).

599. *See* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶ 58 (holding that the data subject's entitlement runs to an explanation of "the procedure and principles actually applied," not disclosure of the underlying model).

or qualified experts can verify the consumer-facing explanation if the borrower challenges the decision.[600]

### *B. Modes of explanation*

A workable right to explanation in automated credit scoring must be built as a layered communication duty, because the legal point is contestability and risk control.[601] This immediately implies two distinct explanatory targets, one of which concerns how the scoring system is designed and what it is for, while the other concerns why this person received this outcome and what could change it.[602] On this basis, a context-sensitive and tiered explanation mode should be employed.[603] Credit scoring involves routine processing and high-stakes adverse decisions, and the informational burden should increase with dispute impact and intensity.[604] One systematized proposal treats many models as resistant to global causal storytelling and therefore prefers indirect explanation routes that disclose stable and testable relations between inputs and outputs.[605] It allocates different explanation intensities to ordinary, key, and contested scenarios.[606] This

600. *See* 12 C.F.R. pt. 1002, Supp. I, § 1002.9, para. 9(b)(2) (requiring disclosed reasons to correspond to the factors actually considered or scored); CFPB 2023, *supra* note 51; U.S. DEP'T OF THE TREASURY, *supra* note 111, at 20–21; EURO. BANKING AUTH., EBA/GL/2020/06, *Guidelines on Loan Origination and Monitoring*, at 54–55, 50–62 (May 29, 2020).

601. *See* Kaminski, *supra* note 36, at 204, 209, 211 (discussing explanation as a governance tool for accountability, contestation, and oversight); *see* Edwards & Veale, *supra* note 37, at 41, 43 (warning that explanation rights must be tied to contestability and institutional function rather than abstract transparency); Kaminski & Malgieri, *supra* note 380, at 127–28, 139–40 (2021).

602. *See* Wachter et al., *supra* note 88 (distinguishing an explanation of "the logic, significance, envisaged consequences and general functionality" of a scoring system from an explanation of "the weighting of features, machine-defined case-specific decision rules, information about reference or profile groups" for a specific outcome); *but cf.* Selbst & Powles, *supra* note 56, at 4–6 (discussing how the GDPR's provisions inherently support a "right to explanation" for automated decisions, emphasizing the importance of meaningful information about the logic involved to ensure transparency, accountability, and fairness in algorithmic processes).

603. *See* Kaminski, *supra* note 36, at 209–11 (describing the GDPR's approach as "qualified transparency" –"a system of targeted revelations of different degrees of depth and scope aimed at different recipients"—rather than a single undifferentiated disclosure standard); *see also* Söderlund, *supra* note 41, at 100–05 (arguing that the EU AI Act's tiered transparency mandates for oversight bodies operationalize qualified transparency in practice).

604. 12 C.F.R. pt. 1002, Supp. I, § 1002.9, para. 9(b)(2); *See* WP29, *supra* note 152, at 21 (explaining that automated decisions trigger heightened safeguards where they "significantly affect the circumstances, behavior or choices" of the individual, have "a prolonged or permanent impact," or "at its most extreme, lead to the exclusion or discrimination of individuals"); *see also* Kaminski, *supra* note 36, at 209 (describing "qualified transparency" as calibrating disclosure depth to context rather than applying a uniform standard).

605. *See* Ribeiro, *supra* note 336, at 1135–37; *see* Wachter et al., *supra* note 356, at 841–45 (proposing unconditional counterfactual explanations that describe "the smallest change to the world that can be made to obtain a desirable outcome" without relying on causal models or opening the black box).

606. *See* Su Yu (苏宇), *Suanfa Jieshi Zhidu de Tixihua Goujian* (算法解释制度的体系化构建) [*The Systematic Construction of Algorithm Explanation System*], 1 Dongfang Faxue (东方法学) [ORIENTAL L. REV.] 91, 91 (2024) (arguing that algorithmic explanation pathways and technical approaches should be calibrated separately for routine, key, and contested scenarios, with explanations fixed through freezing, sampling, and mirroring mechanisms subject to external verification); *cf.* Chuxue Cao et al., *SafeLawBench: Towards Safe Alignment of Large Language Models*, FINDINGS ASS'N COMPUTATIONAL LINGUISTICS: ACL 2025, 14015, 14015–17 (2025) (proposing a legally grounded LLM safety benchmark that classifies safety risks into hierarchical levels based on legal standards, including critical personal safety, property and living security, fundamental rights, and welfare protection; showing how legal standards can operationalize risk-sensitive evaluation rather than relying on undifferentiated or subjective safety categories).

approach reframes explainability as an evidential output that can support review, compliance, and remediation, even when the underlying model is nonlinear.[607] It also provides a proportionality logic that can withstand trade secret objections because the duty is to provide usable external reasons rather than publishing internal technical details.[608] As a result, the next question concerns operationalizing hard and soft modes without allowing soft narratives to replace hard evidence when rights are at stake.[609]

A hard explanation should be treated as an accountability instrument that produces checkable claims about variable influence and decision sensitivity, since that is what allows a reviewer to test error, proxy use, and discriminatory effects.[610] In practice, hard explanations can be delivered through structured feature contribution disclosures, sensitivity-style perturbation tests, and ranked influence bands that translate complex dependence into a form that a regulator or expert can audit and a layperson can still use.[611] A particularly defensible legal output is an explanation that specifies which factual changes would have altered the score, because it avoids speculative storytelling and instead identifies decision-relevant facts that support correction and resubmission.[612] This mode is also easier to police for honesty because it can be assessed against the controller's actual pipeline and logs, which is why it pairs naturally with proposals for algorithm inspection and fixation mechanisms that preserve an objective record for dispute resolution.[613]

607. *See* Ribeiro, *supra* note 336, at 1135–37 (defining SHAP as a unified additive feature-attribution framework in which model linearity is an optional simplifying assumption, so the method applies to nonlinear models); Lundberg & Lee, *supra* note 336 (same).

608. *See* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 68–73 (holding that trade-secret protection cannot justify an absolute refusal to explain, and that contested material should instead be disclosed to a supervisory authority or court for case-specific balancing); *see also* Case C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, ¶¶ 43–46 (Opinion of AG Pikamäe) (reasoning that confidentiality interests may limit the form of disclosure but not extinguish the explanation duty).

609. *See* Liberti & Petersen, *infra* note 618, at 38–41 (distinguishing hard information, verifiable independent of its collector, from soft information).

610. *See* Patrick Rehill & Nicholas Biddle, *Transparency Challenges in Policy Evaluation with Causal Machine Learning: Improving Usability and Accountability*, 6 DATA & POL'Y 43, § 4.1.1. (2024) (discussing how transparency in causal machine learning, particularly in policy evaluation, is crucial for ensuring accountability and usability by addressing the black-box nature of models through explainable AI (XAI) and interpretable AI (IAI) techniques; "Accountability can be seen as a multi-stage process consisting of providing information for investigation, providing an explanation or justification, and facing consequences if needed. The problem with accountability for machine learning systems is clear: they can obfuscate the exact nature of the failure, making it very difficult to obtain an explanation or justification.").

611. *See* Ribeiro, *supra* note 336, at 1135–37 (introducing LIME, which perturbs inputs and fits a local surrogate model to produce per-feature contribution estimates for any classifier); Lundberg & Lee, *supra* note 336 (proposing SHAP, which assigns each feature an additive contribution value to a specific prediction, yielding ranked influence across inputs).

612. *See* Du Xiaoqi (杜小奇), *Duoyuan Xiezuo Kuangjia xia Suanfa de Guizhi* (多元协作框架下算法的规制) [*The Regulation of the Algorithm Under the Multi-Collaborative Framework*], 37 HEBEI FAXUE (河北法学) [HEBEIL.SCI.] 180, 188 (2019) (discussing how a multi-collaborative regulatory framework involving individuals, algorithm deployers, and regulators can enhance algorithmic explainability by focusing on the prevention and correction of adverse algorithmic outcomes rather than merely explaining outputs); *see also* Wachter et al., *supra* note 356, at 841–45 (proposing counterfactual explanations that identify "the smallest change to the world that can be made to obtain a desirable outcome").

613. *See* Su Yu (苏宇), *supra* note 606, at 91 (proposing freezing, sampling, and mirroring mechanisms to fix algorithmic explanations for external verification); *see also* Zhou Xiang (周翔), *Suanfa Ke Jieshi Xing: Yige*

Moreover, soft explanation should be treated as a legally constrained translation layer.[614] This is because its legitimacy lies in making the hard outputs comprehensible rather than in replacing them with persuasive rhetoric.[615] A coherent practice involves expressing the core decision factors in plain language, supported by visual summaries and interactive pathways that let the individual see what data categories mattered and which steps are available for human review or appeal.[616] Soft explanation matters where the credit relationship is ongoing, and the borrower is expected to act on the explanation by correcting data or supplying supplementary evidence.[617] Its communicative value, however, depends on a harder institutional record.[618] The comparative analysis above shows the same problem in different forms.[619] EU law may require the provision of meaningful information without ensuring access to provider-side evidence.[620] Chinese law may recognize processor-side explanation duties, even as platform responses remain one-way, generic, or non-reviewable.[621] U.S. adverse-action law offers a more concrete model for giving reasons, yet factor labels may still leave proxy effects, alternative data, or vendor-controlled scoring untested.[622]

*Jishu Gainian de Guifan Yanjiu Jiazhi* (算法可解释性：一个技术概念的规范研究价值) [*Algorithmic Interpretability: The Normative Research Value of a Technical Concept*], 3 BIJIAO FA YANJIU (比较法学研究) [J.COMP.L.] 180, 196 (2023) (discussing how the technical concept of algorithmic explainability, encompassing both global model transparency and localized output justification, is crucial for accurately identifying and mitigating algorithmic risks).

614. *See* Liberti & Petersen, *infra* note 618, at 38–41 (distinguishing hard information, transmissible without loss of content, from soft information, which cannot be separated from the context that produced it).

615. *See* Bello et al., *supra* note 339, at 12–22 (proposing natural-language mediation that translates technical explanation outputs for non-technical stakeholders without substituting for them); Wachter et al., *supra* note 356, at 871–78.

616. *See* WP29, *supra* note 152, at 31–32 (recommending visualization and interactive techniques, and mechanisms allowing data subjects to check their profile and the sources used to develop it); *see also* Sirui Han et al., *Reimagining Legal Fact Verification with GenAI: Toward Effective Human-AI Collaboration,* PROC. 2026 CHI CONF. ON HUM. FACTORS IN COMPUTING SYS., APR. 2026, at 9–10 (finding that lawyers expect AI outputs to be traceable through sources, links, screenshots, and layered formats such as reports or mind maps; noting that "Each conclusion should be marked with its source" and that users should be able to "inspect AI reasoning step by step"); *see also* Zheng Zhihang (郑智航) & Xu Zhaoxi (徐昭曦), *Dashuju Shidai Suanfa Qishi de Falü Guizhi yu Sifa Shencha—-Yi Meiguo Falü Shijian wei Li* (大数据时代算法歧视的法律规制与司法审查——以美国法律实践为例) [*Legal Regulation and Judicial Review of Algorithmic Discrimination in the Age of Big Data: Take Legal Practice in the U.S. as an Example*], 4 BIJIAO FA YANJIU (比较法学研究) [J. COMP. L.] 111, 121 (2019) (discussing how legal regulations and judicial reviews address algorithmic discrimination by implementing principles of transparency, accountability, and fairness).

617. CFPB 2022, at 3; *see* Wachter et al., *supra* note 356, at 841–45 (framing explanation as a means of enabling the data subject to act and alter future outcomes, not merely to understand a past decision).

618. *See* José María Liberti & Mitchell A. Petersen, *Information: Hard and Soft*, 8 REV. CORP. FIN. STUD. 1, 38–41 (2019) (distinguishing hard information, which is quantitative and transmissible without loss of content, from soft information, which loses content when reduced to a number).

619. *See supra* notes 539–47 and accompanying text; *see also* Li & Chen, *supra* note 57 (showing that formally similar data-protection language acquires divergent regulatory meaning across jurisdictions).

620. *See* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 58, 64–65 (requiring disclosure of applied procedures and principles without conferring entitlement to provider-side records); *see also* Edwards & Veale, *supra* note 37, at 43; GDPR, art. 15(1)(h).

621. *See* PIPL, art. 24(3) (entitling individuals to request an explanation and refuse solely automated decisions); *see also* Xu Yong v. Zhima Credit Mgmt. Co., Hangzhou Internet Ct., (2018) Zhe 0192 Min Chu No. 302 (China) (illustrating generic platform responses providing no reviewable basis for challenge).

622. *See* 12 C.F.R. pt. 1002, Supp. I, para 9(b)(2)–(3) (providing that a creditor "need not describe how or why a factor adversely affected an applicant"); *see also* Hurley & Adebayo, *supra* note 16, at 180–87 (arguing reason codes leave proxy effects and alternative-data inputs untested).

Explanation modes in credit scoring should therefore combine a consumer-facing account with a reviewable evidential layer.[623] Fair-lending work on algorithmic fairness reinforces this design.[624] Sufficiency, separation, and equality of opportunity reflect different normative views of fair credit allocation, may conflict with one another, and may preserve bias where observed data already encode past discrimination.[625] A defensible mode of explanation should pair an outward adverse-action account with preserved internal materials on validation, data lineage, subgroup performance, and reason mapping, so that soft explanation remains understandable while hard explanation remains verifiable.[626]

### *C. Gradient requirements*

The decisive question is not whether an algorithm is explainable in an abstract technical sense but rather what level of explanation the law should require for a particular decision context, given the structural information asymmetry between controllers and affected persons and the practical need for contestability.[627] A content gradient, anchored in proportionality but operationalized through intelligibility and remedy, offers a legally defensible middle path

623. *See* Kaminski, *supra* note 36, at 209–16 (describing "qualified transparency" as pairing individual-facing disclosure with deeper systemic transparency to regulators and auditors).

624. *See* Rashed Bahlool et al., *Performance, Fairness, and Explainability in AI-Based Credit Scoring: A Systematic Literature Review*, 19 J. RISK FIN. MGMT. at 1, 3, 13–14 (2026) (surveying recent AI-based credit-scoring literature and finding that performance, fairness, and explainability are often treated separately, while responsible deployment requires their joint treatment under regulatory oversight and human review).

625. *See* I. Elizabeth Kumar et al., *Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with U.S. Fair Lending Regulation*, PROC. 2022 AAAI/ACM CONF. ON AI, ETHICS, & SOC'Y 357, 357–68 (2022) (mapping computational fairness definitions onto ECOA and Regulation B frameworks); Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 BIG DATA, 153, 153–56 (2017); Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV., 812, 831–45 (2020) (arguing competing fairness measures embed incompatible moral commitments).

626. *See* EUROPEAN BANKING AUTH., EBA/GL/2020/06, GUIDELINES ON LOAN ORIGINATION AND MONITORING, EBA/GL/2020/06, paras. 54–55, 60–62 (May 29, 2020), https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring [https://perma.cc/SSG4-JKMF] (requiring traceability and auditability of automated model inputs and outputs and data infrastructure enabling effective audit trailing of rating or scoring overrides); *see also* Weijie Shi et al., *LegalReasoner: Step-Wised Verification-Correction for Legal Judgment Reasoning*, 1 PROC. 63RD ANN. MEETING ASSOC. FOR COMPUTATIONAL LINGUISTICS 7297, 7297–7304 (2025) (proposing a legal judgment reasoning framework that decomposes complex cases into dispute points, verifies each reasoning step through a process verifier, and applies expert-designed attribution and correction strategies when reasoning errors are detected); *see also* Kaminski, *supra* note 36, at 214–16; *cf.* Haoran Li et al., *ContextLens: Modeling Imperfect Privacy and Safety Context for Legal Compliance*, 1 PROC. 64TH ANN. MEETING ASS'N FOR COMPUTATIONAL LINGUISTICS 6503, 6503–05 (2026) (proposing a semi-rule-based framework that identifies known, unknown, and ambiguous contextual factors in LLM compliance assessments involving the GDPR and EU AI Act); *cf.* Yujin Zhou et al., *LexAgentHallu: A Hierarchical Benchmark for Profiling Hallucinations in Legal Agents*, PROC. 2026 CONF. ON EMPIRICAL METHODS IN NATURAL LANG. PROCESSING 7, 16 (2026) (using full agent trajectories to identify hallucinations despite correct final answers, illustrating the value of preserving reasoning and execution records for verification).

627. *See* Tim Miller, *Explanation in Artificial Intelligence: Insights from the Social Sciences*, 267 A.I. 1, 4 (2019) (arguing that the field of explainable artificial intelligence (XAI) can leverage existing research in philosophy, psychology, and cognitive science to develop more effective and intuitive explanations, addressing cognitive biases and social expectations inherent in the explanation process).

because it ties the depth of explanation to both the degree of automation and the degree of rights impact.[628]

### *1. Proportionality*

Some scholarship proposes that algorithmic explanation should be structured by a pair of principles, namely, substantive explanation and proportionality, and that substantive explanation should extend to effects on fair competition order and market transparency;[629] however, this expansion has not generated broad doctrinal convergence, suggesting that the more stable normative anchor is proportionality.[630] Proportionality matters because explanation is an informational right and a governance technique and must distribute compliance burdens in a manner that preserves innovation without sacrificing effective rights protection.[631] The EU AI Act makes this governance logic explicit by providing that, where a person is affected by a decision based on data by a high-risk AI system, they are entitled to obtain from the deployer "*clear and meaningful explanations of the role of the AI system in the decision-making procedure*" under Article 86(1).[632] That clause should be read as setting a minimum functional threshold, rather than as mandating the disclosure of the model itself, because EU data protection law already treats intelligibility as the normative yardstick, not the disclosure of source code.[633]

### *2. Trigger thresholds*

A workable gradient begins with disciplined triggering conditions.[634] In Chinese law, Article 24(3) of the PIPL does not establish explicit criteria for

628. *See* Hulianwang Xinxi Fuwu Suanfa Tuijian Guanli Guiding (互联网信息服务算法推荐管理规定) [Regulations on the Administration of Algorithmic Recommendations in Internet Information Services] (promulgated by the Cyberspace Admin. of China et al., Dec. 31, 2021, effective Mar. 1, 2022) arts. 12, 17 (China) (requiring disclosure of the basic principles, purposes, and operational mechanisms of recommendation algorithms in an appropriate manner); *see also* Su Yu (苏宇), *supra* note 606, at 91 (calibrating explanation pathways separately for routine, key, and contested scenarios).

629. Zhang Xin (张欣), *Suanfa Jieshi Quan yu Suanfa Zhili Lujing Yanjiu* (算法解释权与算法治理路径研究) [*Research on the Right to Algorithmic Explanation and Algorithmic Governance Paths*], 6 Zhong Wai Faxue (中外法学) [PEKING U. L.J.] 1425 (2019).

630. *See* Brkan & Bonnet, *supra* note 346, at 23–28 (surveying divergent legal and technical approaches without identifying doctrinal convergence).

631. *See id.* at 23 (discussing how the legal and technical feasibility of the GDPR's right to explanation for algorithmic decisions involves navigating complexities such as algorithmic secrecy, technical obstacles, and the need for explanations to be both understandable to non-experts and legally compliant).

632. EU AI Act, art. 86(1) ("Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, with the exception of systems listed under point 2 thereof, and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken").

633. *See* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶ 58; *see also* EU AI Act, Recital 171 (describing the Article 86 right as covering the AI system's role and the main elements of the decision, not model disclosure).

634. *See* WP29, *supra* note 152, at 21–22 (setting out threshold criteria for when automated processing produces legal or similarly significant effects); GDPR, arts. 15(1)(h), 22 (1), 2016 O.J. (L 119) 43, 46.

determining what constitutes a major impact.[635] The legislative choice of vagueness is plausibly intentional given the diversity of automated decision settings.[636] In a credit-scoring context, the most defensible operational test is whether the score alters the legal position or economic interests through contracting leverage, not merely through informational classification.[637] Typically, the formation or termination of a contract should qualify as an adverse state within the meaning of that provision.[638] If an individual cannot sign a contract because of a low score, they should be treated as having suffered a significant impact.[639] Meanwhile, if an ongoing contract is terminated because of the scoring output, the scored person should be able to invoke the refusal and challenge mechanisms.[640] Crucially, significant impact should not be limited to outright denial, because coercion can be implemented through degraded terms that frustrate contractual purpose while preserving formal assent.[641] If a person can only contract under materially unfavorable conditions, their legal position is adversely affected;[642] for instance, where they are denied cash on delivery or prevented from using credit card payment options that others may use.[643] This functional reading is reinforced by the Chinese normative matrix, because the PIPL expressly foregrounds the automated processing of sensitive information in contexts such as transaction pricing, commercial marketing, information pushing, and credit assessment, which are precisely the settings in which automated outputs can crystallize into rights-relevant consequences.[644]

635. *See* PIPL, art. 24(3); *see also* Wang Yegang (王叶刚), *Geren Xinxi Chulizhe Suanfa Zidonghua Juece Zhihai de Minshi Zeren* (个人信息处理者算法自动化决策致害的民事责任) [*Civil Liability for Harm Caused by Algorithmic Automated Decision-Making by Personal Information Processors*], Zhongguo Renmin Daxue Xuebao (中国人民大学学报) [J. RENMIN UNIV. CHINA] (2022) (noting Article 24 states automated decision-making rules without specifying criteria or consequences of breach).

636. *See* Ding et al., *supra* note 58 (describing China's risk-based approach as deliberately open-textured to accommodate varied processing contexts).

637. *See* Case C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, ¶¶ 48–50 (treating a probability value as an automated decision where a third party draws strongly on it to establish, implement, or terminate a contractual relationship).

638. *See SCHUFA Holding (Scoring)*, ECLI:EU:C:2023:957, ¶ 50; *see also* WP29, *supra* note 152, at 21 (listing cancellation of a contract as a legal effect).

639. *See* WP29, *supra* note 152, at 22 (identifying decisions affecting financial circumstances and access to credit); *see also* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 20–22 (credit assessment resulting in refusal of a mobile phone contract).

640. *See* PIPL art. 24(3) (right to request explanation and to refuse decisions made solely by automated decision-making).

641. *See* WP29, *supra* note 152, at 22 (explaining that differential pricing triggers Article 22 where "prohibitively high prices effectively bar someone from certain goods or services").

642. *See* PIPL, art. 24 (prohibiting unreasonable differential treatment in transaction conditions such as price); *see also* WP29, *supra* note 152, at 22.

643. *See Tao Bao Xin Yong Fen 1 Fen Ke Yi Fa Huo Ma ？* (淘宝信用分1分可以发货吗？)[*Can Goods Be Shipped with a Taobao Credit Score of 1?*], TAOBAO HELP CTR., https://bk.taobao.com/k/taobaoxinyong_510/8b6b26c46d6c551c8bb2656c744628c7.html [https://perma.cc/TT76-6NYJ] (last visited Mar. 10, 2025) (indicating that low-scoring users lose access to cash-on-delivery and certain payment options available to others).

644. *See* Xue Hong (薛虹), *Lun Dianzi Shangwu Hetong Zidong Xinxi Xitong de Falu Xiaoli* (论电子商务合同自动信息系统的法律效力) [*On the Legal Effect of Automatic Information Systems in E-commerce Contracts*], 1 SUZHOU DAXUE XUEBAO (苏州大学学报(哲学社会科学版)) [J.SOOCHOW UNIV.(PHILOS.SOC.SCI.ED.)] 70 (2019) (discussing how the legal framework established by the Electronic Commerce Law in China acknowledges the legal effect of automatic information systems in e-commerce contracts to safeguard the rights and interests of all parties involved and ensure the reliability of e-commerce transactions).

A parallel significance threshold exists in the GDPR.[645] Article 22 is narrower than it first appears because it only applies to automated decisions that (i) produce legal effects for the data subject or (ii) similarly and significantly affect them.[646] The content of the latter limb is not quite clear, considering what is already stated in the former;[647] therefore, the practical threshold has to be stabilized through guidance rather than textual inference.[648] However, as suggested by the WP29, "similar effect" can be interpreted as significantly affecting the circumstances, behavior, or choices of data subjects.[649] These decisions include those that affect a person's financial situation, including their eligibility for credit; prevent a person from accessing a job opportunity or place them at a serious disadvantage; and affect a person's access to education, such as admission to higher-education institutions.[650] These aspects affect the gradient design because if the trigger remains indeterminate, controllers can argue that a credit score merely informs a commercial preference, whereas the subject experiences it as a gatekeeping constraint with life-shaping consequences.[651] Accordingly, the impact limb should be read functionally in terms of whether the decision plausibly reshapes the person's feasible set of choices.[652]

### *3. The content gradient*

The core doctrinal move at the content stage is to shift attention from explaining the AI as an artifact to explaining the AI-supported decision as a contestable act of governance.[653] A person does not need a component-by-component decomposition of the model to exercise rights effectively, but they do need an interaction-friendly account of what drove this outcome and what can be corrected.[654] Miller's social science account makes this link operational by explaining why usable explanations are contrastive, selective, causal in aspiration rather than purely correlational, and dialogic in form, which means that they must be tailored to the recipient's background assumptions to enable genuine

645. *See generally*, GDPR, art. 22, at 1, 39 (highlighting the significance threshold in the GDPR).

646. *Id.*

647. Opinion of Advocate General Pikamäe, Case C-634/21, SCHUFA Holding AG, ECLI:EU:C:2023:220, ¶ 49 (Mar. 16, 2023).

648. *See* Luca, *supra* note 150 (highlighting suggested guidance for interpreting Article 22(1)).

649. *See* WP29, *supra* note 152, at 21 (discussing how similar effect can affect the circumstances).

650. *See id.* ("[f]or data processing to significantly affect someone the effects of the processing must be sufficiently great or important to be worthy of attention. In other words, the decision must have the potential to: significantly affect the circumstances, behavior or choices of the individuals concerned; have a prolonged or permanent impact on the data subject; or at its most extreme, lead to the exclusion or discrimination of individuals.").

651. *See id*. at 22 (discussing the potential life-shaping consequences of being denied a credit score).

652. *See SCHUFA Holding AG*, ECLI:EU:C:2023:220, ¶43 (explaining that in order to align with the intent of the regulation, 'decision' should be strictly interpreted, but instead refer to the effect the score has on the subject).

653. *See generally*, Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 54–56 (highlighting the AI-decision as a contestable decision).

654. *Id.* ¶¶ 54–56, 59–62.

understanding.[655] This human-centered logic strengthens the gradient approach.[656] It provides a principled reason for rejecting both extremes, dumping technical artifacts that users cannot parse and offering vague assurances that users cannot dispute.[657] A defensible gradient should be built around three cumulative benchmarks—namely intelligibility, credibility, and remedy.[658] The CJEU's interpretation of Article 15(1)(h) in C-203/22 frames "meaningful information" as functional, important, relevant, and intelligible while defining "logic involved" as information about the "procedure and the principles" that lead to a specific result, which sets the doctrinal content baseline for post-decision explanation.[659]

First, the pre-scoring explanation is justified because it enables behavioral guidance and informed engagement with the scoring system.[660] At this tier, the controller should disclose the purpose and intended use of scoring, categories of inputs, the existence of profiling and inference, potential outcomes, and the main rights and channels for redress.[661] Furthermore, where sensitive information is used, the explanation must identify sensitive categories and protective measures, given the heightened duties under Article 28 of the PIPL and the potential harm from the misuse of financial and behavioral data.[662] If data are to be shared or transferred, the recipients and purposes must be disclosed under Articles 22 and 23 of the PIPL;[663] meanwhile, cross-border transfers require additional disclosures under Article 39 because the risk profile changes materially once credit data circulate.[664]

Second, the post-scoring explanation must be specific to the affected person and the challenged decision, given that its purpose is to enable the person to understand it, express a view, and contest the outcome, as required by the functional link between access rights and safeguards under Article 22(3) of the GDPR.[665] At a minimum, the controller should provide a sufficiently detailed account of the method used to calculate the score and the reasons for this result, along with general information regarding the factors considered and their relative weights at an aggregate level;[666] this is the least intrusive way to make the decision contestable without forcing full disclosure of the algorithm.[667] Moreover, this tier should reflect how humans understand explanations, where controllers should provide contrastive and selective reasons and articulate causal

655. *See* Miller, *supra* note 627, at 3.
656. *See generally*, *id.* at 1–3 (describing the need for human-centric logic).
657. *Id.*
658. *Id.*
659. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 38–42.
660. *See* PIPL, art. 17 (requiring personal information processors to give a pre-scoring explanation).
661. *Id.*
662. PIPL, arts. 28–30.
663. PIPL, arts. 22–23.
664. PIPL, art. 39.
665. GDPR, art. 22(3).
666. Opinion of Advocate General Pikamäe, Case C-634/21, SCHUFA Holding AG, ECLI:EU:C:2023:220, ¶ 49 (Mar. 16, 2023), ¶ 58.
667. *Id*. ¶¶ 56–58.

pathways rather than mere correlations.[668] In credit scoring, this implies disclosing the decisive feature groups and their direction of effect, indicating the main adverse drivers, and supplying a counterfactual improvement route where appropriate.[669]

Third, where the person plausibly alleges error or unfairness, the controller must provide a correction-oriented explanation identifying whether the issue lies in data accuracy, feature construction, model behavior, or deployment constraints.[670] Without this information, the right to explanation degenerates into a right to be told rather than one to be protected.[671] This tier must culminate in a decision pathway, such as re-scoring with corrected data, reassessment with meaningful human review, or suspension of automated reliance, because this remedy is the legal point of explanation and not an optional add-on.[672] Under Article 24(3) of the PIPL, the user's right to refuse decisions made solely through automated decision-making reinforces the need for an exit or override mechanism in high-impact credit contexts.[673]

### *4. Trade secrets and third-party data*

The law treats commercial confidentiality as a calibration factor, not an on–off switch that extinguishes the essence of access and contestation rights.[674] Recital 63 of the GDPR rejects absolute refusal and requires that at least a minimum amount of information must be provided that preserves the essence of the right to data protection.[675] This means that the minimum disclosure standard must be

668. *See* Miller, *supra* note 627, at 3 (explaining why controllers should provide contrasted and selected reasons).

669. Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 55–60.

670. *Id.* ¶¶ 61–63.

671. *Id.* ¶¶ 38–42.

672. GDPR, art. 22(3) ("In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision"); *Dun*, ECLI:EU:C:2025:117, ¶¶ 54–56, 58–62, 65 (holding that the right to obtain meaningful information must enable the data subject to understand the reasons for an automated outcome, verify whether errors or unfairness are present, and effectively exercise corrective and contestatory rights under Article 22(3) GDPR); cf. Sirui Han et al., *Hong Kong Generative Artificial Intelligence Technical and Application Guideline*, DIGITAL POLICY OFFICE, HKSAR GOV'T § 2.2.2, at 42 (2025), https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/doc/HK_Generative_AI_Technical_and_Application_Guideline_en.pdf [https://perma.cc/K8UY-V792] (discussing Hong Kong banking guidance on opt-out and human-intervention options, as far as practicable during early deployment of customer-facing GenAI applications, and review channels where opt-out is unavailable).

673. *See* Xiao Mengli (肖梦黎), *Pingtai Xinyong Pingfen Jizhi de Guizhi Gongneng jiqi Fazhihua Shixian* (平台信用评分机制的规制功能及其法治化实现) [*The Governance and Legalization of Platform Ranking Mechanism*]), 2 TANSUO YU ZHENGMING (探索与争鸣) [EXPLOR. FREEVIEWS.] 133, 141 (2022) (discussing how the platform credit scoring mechanism constructs a new scoring society with far-reaching impacts on legal patterns and trust systems, emphasizing the necessity for explainability and transparency in scoring algorithms to mitigate regulatory risks such as non-standardization, discrimination, and value misalignment).

674. *See* Mark-Silas A. Malekela, *AI and Confidentiality Protection in International Commercial Arbitration: Analysis of the Existing Legal Framework*, 5 DISCOVER A. I. 83 (2025) (discussing how the integration of Artificial Intelligence (AI) in international commercial arbitration raises critical concerns regarding confidentiality, as AI tools may inadvertently disclose sensitive information, necessitating robust legal frameworks and procedural safeguards).

675. GDPR, recital 63, 2016 O.J. (L 119) 1, 12.

defined in functional terms rather than by reference to the controller's preferred frame of secrecy.[676] A more analytically robust account of informational rights situates access within an explicit balancing architecture between Articles 8 and 17(2) of the EU Common Financial Reporting framework, instead of treating transparency as a self-contained slogan.[677] Based on that approach, conflicts with trade secrets, intellectual property, and software copyrights cannot justify an absolute refusal to provide information, because some minimum disclosure is required to ensure that the essence of the right to the protection of personal data is not compromised.[678] Article 12 provides the operative translation of this balance, imposing an intelligibility standard that is breached by secrecy and performative transparency.[679] Information must be concise, transparent, intelligible, and easily accessible, using clear, plain language; moreover, disclosing a complex formula without contextual clarification counts as obfuscation.[680] In alignment with that logic, Articles 12 and 15 exclude any general duty to disclose the algorithm as such yet also require sufficiently detailed explanations of the method used to calculate the score and the reasons for a particular outcome, along with general information regarding the factors considered and their relative aggregate-level weighting.[681] Thus, the decision becomes challengeable without collapsing into full model disclosure.[682] This distinction has regulatory power because controllers routinely invoke trade secrets as a general justification to resist access requests, including in high-stakes labor-allocation settings where opacity directly entrenches bargaining asymmetry.[683]

When user-facing disclosure reaches its legitimate limit because of third-party personal data or protected secrets, the governance architecture must shift from individual access to institutional inspection, such that rights effectiveness is preserved through supervised balancing rather than abandoned through opacity.[684] A normatively attractive approach distinguishes disclosure of the algorithm from that of the method and the reasons for the result.[685] In the same vein, where effective challenge depends on information that implicates third-party

676. *Id.*

677. Opinion of Advocate General Pikamäe, Case C-634/21, SCHUFA Holding AG, ECLI:EU:C:2023:220, ¶ 49 (Mar. 16, 2023), ¶ 55.

678. *Id.*¶ 56.

679. *Id.*¶ 557.

680. *See* Mariavittoria Masotina & Anna Spagnolli, *Transparency of Privacy Notices and Contextualisation: Effectively Conveying Information Without Words*, 41 BEHAV. & INFO. TECH. 2120, 2121–22 (2022) (discussing how the comprehensibility of privacy notices is influenced by the sequential context in which they appear, highlighting that the action preceding the notice affects visitors' understanding of its cause and content, rather than the explicit wording of the notice itself).

681. *See* GDPR, art. 12, 15 (discussing the requirements of the GDPR for the algorithm and weighting).

682. *SCHUFA Holding AG*, ECLI:EU:C:2023:220, ¶ 57.

683. *See* Worker Info. Exch. & App Drivers & Couriers Union v. Uber & Ola, ECLI:NL:GHAMS:2023:54, ¶¶ 13–15 (Neth.) (holding Uber cannot rely on the protection of trade secrets as a general ground for refusing access. . .particularly where the requested information concerns the allocation of work and remuneration, which directly affects the position of drivers).

684. Charter of Fundamental Rights of the European Union, art. 47, 2012 O.J. (C 326) 391 ("Everyone whose rights and freedoms guaranteed by the law of the Union are violated has the right to an effective remedy before a tribunal in compliance with the conditions laid down in this Article.").

685. *SCHUFA Holding AG*, ECLI:EU:C:2023:220, ¶ 58.

data or confidentiality, the CJEU accepts that disclosure to a supervisory authority or court may be required, with the balancing performed institutionally.[686] For credit scoring, this results in a two-track design: a user-facing explanation that is intelligible and sufficiently detailed for contestation along with a regulator- or court-facing dossier that can be inspected under confidentiality safeguards when the dispute turns to deeper technical or comparative claims.[687]

### *D. Procedural design for disclosure*

The duty of case elucidation exists precisely because litigation can become unfair when one party controls the relevant facts and the other cannot realistically establish causation or fault.[688] Within this logic, the right to explanation should be treated as a procedural supplement that lowers the threshold for activation.[689] The duty of case elucidation remains conditioned on an initial plausible factual basis, the absence of subjective fault, and the presence of important, specific, and reasonably foreseeable facts.[690] However, automated scoring routinely frustrates these conditions because the claimant is confronted with a black box, continuous updates, and institutionalized refusal justified by technical complexity or trade secrets.[691] The standard should, therefore, not demand early technical particulars that only the defendant can access;[692] rather, it should require a credible demonstration that the claimant has suffered a concrete adverse outcome in a credit context and that the defendant is monopolizing the decisive

686. Case C-268/21, Norra Stockholm Bygg AB v. Per Nycander AB, ECLI:EU:C:2023:145, ¶ 58.

687. *SCHUFA Holding AG*, ECLI:EU:C:2023:220, ¶ 58; Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶ 76.

688. *See, e.g.*, Fed. R. Civ. P. 56 (highlighting U.S. law establishing what to do when one party cannot establish causation).

689. Selbst & Powles, *supra* note 56.

690. *See* Pang Xiaoju (庞小菊) & Xu Yingqian (徐英倩), *Bu Fu Zhengming Zeren Dangshiren de Shian Jieming Yiwu* (不负证明责任当事人的事案解明义务) [*On the Obligation to Explain Cases of Litigant Without Burden of Proof*], 23 GUOJIA JIANCHAGUAN XUEYUAN XUEBAO (国家检察官学院学报) [J. NAT'L PROSECUTORS COLL.], 115–30 (2015) (discussing how the duty to explain cases in civil litigation, particularly the obligation of parties who do not bear the burden of proof to clarify relevant facts aid in the discovery of truth and ensuring fairness) [hereinafter Pang].

691. *See* Xu Feng (徐风), *Rengong Zhineng Suanfa Heixiang de Falu Guizhi—Yi Zhineng Tougu Wei Li Zhankai* (人工智能算法黑箱的法律规制——以智能投顾为例展开) [*Legal Regulation of AI Algorithm Black Boxes: An Analysis Using Robo-Advisors as an Example*], 6 DONGFANG FAXUE (东方法学) [ORIENTAL L. REV.] 78, 85 (2019) (explaining that AI algorithms are opaque "black boxes" whose hidden-layer processes cannot be fully understood, and that institutions justify non-disclosure on the grounds that algorithms are trade secrets and involve specialized technical language incomprehensible to non-expert consumers even if disclosed).

692. *Cf.* Pan Fangfang (潘芳芳), *Suanfa Qishi de Minshi Zeren Xingtai* (算法歧视的民事责任形态) [*The Forms of Civil Liability for Algorithmic Discrimination*], HUADONG ZHENGFA DAXUE XUEBAO (华东政法大学学报) [ECUPL J.], at 55, 61 (2021) (arguing that once a plaintiff shows a disparate algorithmic impact, the burden shifts to the defendant to prove absence of fault, without requiring the plaintiff to first establish the algorithm's technical operation).

information.[693] This shift follows the same fairness rationale underpinning procedural legibility duties in data protection law.[694]

Comparative procedural design reinforces that evidence disclosure in technologically complex disputes is not an exceptional favor but an institutional response to structural information deficits.[695] The US Federal Rules of Civil Procedure embed discovery with a robust pretrial phase, including the disclosure of electronically stored information, such that factual development does not depend on the weaker party's assumptions.[696] By contrast, Germany and Japan rely on court-managed production orders that compel parties or third parties to produce electronic evidence, thereby operationalizing procedural equality within civil law.[697] These models demonstrate that the trigger for disclosure is not a demand to decode source code but a demonstration that the requested material is necessary to litigate fairly and proportionately.[698] This observation is consistent with the view that procedural equality is constitutionally significant rather than a mere formal exercise, and that the judiciary must be able to neutralize evidentiary monopolization to secure a fair hearing.[699]

693. *See* Case C-203/22, CK v Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117, ¶¶ 16, 18 (CK was refused a mobile telephone contract "on the ground that, according to an automated credit assessment, carried out by D & B, she did not have sufficient financial creditworthiness," and D & B "claim[ed], in essence, that, due to a protected trade secret, it did not have to disclose to CK any information in addition to the information that had already been provided to her.").

694. *See* Casey et al., *supra* note 87, at 151 (explaining that the GDPR's "right to explanation" "mandates that entities handling the personal data of EU citizens 'ensure fair and transparent processing,'" which "requires providing citizens with access to 'meaningful information about the logic involved' in certain automated decision-making systems.").

695. *See* Diego A. Zambrano, *Discovery as Regulation*, 119 MICH. L. REV. 71, 93 (2020) (explaining that the framers of the Federal Rules of Civil Procedure designed broad pretrial discovery to make the pretrial stage "the fundamental information-exchanging event" of litigation, rather than confining information exchange to an exceptional, trial-only occurrence).

696. Fed. R. Civ. P. 26(b)(1) (permitting discovery of any nonprivileged matter relevant to a claim or defense, proportional to the needs of the case, "considering. . . the parties' relative access to relevant information"); Fed. R. Civ. P. 34(a)(1)(A) (permitting requests to produce "electronically stored information. . . stored in any medium from which information can be obtained").

697. Zivilprozessordnung [ZPO] [Code of Civil Procedure], § 142(1) (Ger.) ("The court may order a party or a third party to produce the documents and other records in their possession to which a party has referred."); ZPO § 427 ("If the opposing party fails to comply with the order to produce the document. . . the assertions of the party bearing the burden of proof regarding the nature and content of the document may be considered proven."); Craig P. Wagnild, *Civil Law Discovery in Japan: A Comparison of Japanese and U.S. Methods of Evidence Collection in Civil Litigation*, 3 ASIAN-PAC. L. & POL'Y J. 1, 10–12 (2002) ("A major tool for collecting evidence from an opposing party or third person in Japan is to request that the court order a production of documents," a mechanism "of a compulsory nature" that, following the 1998 amendments, imposes "a general duty on corporations and private citizens to produce the documents that they possess.").

698. *See* Fed. R. Civ. P. 26(b)(1) (limiting discovery to matters "proportional to the needs of the case"); Wachter et al., *supra* note 356, at 843 (arguing that "explanations can, in principle, be offered without opening the 'black box'" of the underlying algorithm).

699. *See* ZHANG WEIPING (张卫平) & XU KE (许可), HIROSHI TAKAHASHI, MINSHI SUSONG FA ZHONGDIAN JIANGYI (民事诉讼法重点讲义) [KEY LECTURES ON CIVIL PROCEDURE LAW] 52–54 ( (TRANS., LAW PRESS 2007) (arguing that procedural equality requires courts to counteract structural evidentiary imbalance; "The so-called procedural equality of the parties does not entail formal symmetry, but rather requires that within the specific structure of litigation, no party should be placed at a substantive disadvantage due to procedural design. . .when evidence or information is effectively monopolised by one party, should the court still adopt a passive stance, the notion of a fair hearing would become a mere empty phrase").

Operationally, China's existing procedural tools can be reorganized into a clearer trigger sequence that moves from targeted requests to coercive consequences, instead of leaving claimants trapped at the pleading stage.[700] Articles 110 and 112 of the *Interpretation by the Supreme People's Court on the Application of the Civil Procedure Law of the PRC* incorporate elements of the duty of case elucidation,[701] and Article 95 of the *Provisions by the Supreme People's Court on Evidence in Civil Procedures* permits adverse presumptions against parties who refuse to disclose evidence without justification after judicial orders.[702] However, as the Pang and Xu conditions are poorly suited to black-box scoring, the court should treat a credible demonstration of adverse credit outcome along with the platform's control of the decisive information as a sufficient condition to activate an order for disclosing decision reasons, data categories, and key influencing variables, with confidentiality measures where appropriate.[703] This logic is illustrated by *Liu Shuxing v. Beijing Bank Credit Card Centre*[704], where a score-driven credit card freeze was treated as a concrete harm that justified judicial intervention once a plausible linkage could be shown.[705] If the institution refuses disclosure, Article 95 should not remain symbolic.[706] Where necessary, courts can adopt a probative exploration mechanism to allow claimants to build plausibility before finalizing the allocation of fault and causation, drawing on analogous evidentiary doctrines in environmental and medical liability.[707]

700. *See supra* notes 697–98 and accompanying text (illustrating this two-step sequence, a targeted production order followed by a coercive consequence for noncompliance, as already operative in the German model under §§ 142, 427 ZPO).

701. Interpretation of the Supreme People's Court on the Application of the Civil Procedure Law of the People's Republic of China (最高人民法院关于适用《中华人民共和国民事诉讼法》的解释), art. 110 (providing that a court may compel a party to appear for questioning on case facts, and may decline to recognize that party's factual claims if the party refuses to appear, be questioned, or sign a truthfulness undertaking, absent other supporting evidence); art. 112 (providing that where documentary evidence is controlled by the opposing party, the court may order its production, and may deem the requesting party's characterization of the evidence's content to be true if the opposing party refuses production without valid reason).

702. Provisions of the Supreme People's Court on Evidence in Civil Litigation (最高人民法院关于民事诉讼证据的若干规定) (2019 amendment), art. 95 (China).

703. *See* Pang Xiaoju (庞小菊) & Xu Yingqian (徐英倩), *supra* note 690, at 115–30 (discussing the conditions governing the duty to explain cases in civil litigation, including an initial plausible factual basis, absence of subjective fault, and the presence of important, specific, and reasonably foreseeable facts).

704. *See* Liu Shuxing v. Beijing Bank Co., Ltd. Credit Card Ctr. (刘树星诉北京银行股份有限公司信用卡中心), (2022) Jing 0113 Min Chu No. 2067, Shunyi Dist. People's Ct. of Beijing (China) (recognizing that the suspension of the claimant's credit card constituted a concrete adverse effect based on the bank's risk assessment, while accepting the bank's reliance on contractual clauses and risk-control needs without requiring disclosure of the underlying scoring logic).

705. *Id*.

706. *See supra* note 702 and accompanying text (permitting adverse presumptions against parties who refuse to disclose evidence without justification after judicial orders).

707. *See* Hu Xuejun (胡学军), *Yongbao yi huo Juchi: Mosuo Zhengminglun de Zhongguo Jingyu* (拥抱抑或拒斥:摸索证明论的中国境遇) [*Embrace or Rejection: The Chinese Situation of the Theory of Probing Proof*], 5 DONGFANG FAXUE (东方法学) [ORIENTAL L.REV.] 76, 78 (2014) (discussing how the concept of "heuristic proof" in civil litigation addresses the challenge of evidence asymmetry by allowing parties with difficulty in proving facts to seek information through court-assisted discovery; "In modern risk-distribution disputes, courts may need to permit a form of probative exploration that allows claimants to establish plausibility before the final allocation of fault and causation, particularly in evidentiary-asymmetric settings").

## VII. CONCLUSION AND FUTURE WORK

The foundation of credit scoring can be traced back to the mid-20th century, when lenders began using manual, rule-based systems to assess creditworthiness, producing numerical scores that guided risk selection.[708] In AI-governed credit markets, automated credit scoring is not simply a technical convenience;[709] rather, it is a decision infrastructure that reallocates opportunity and risk through data-driven inferences while systematically concentrating evidential control in the hands of controllers and deployers.[710] The core legal difficulty is therefore institutional rather than epistemic.[711] The question is how law can make contestability real for the credit consumer when the decisive premises, including data selection, proxy construction, weighting, and operational thresholds, are routinely unavailable to the party who bears the consequences.[712]

This article argues that the right to algorithmic explanation in credit scoring should be understood as a hybrid consumer-protection guarantee.[713] Comparative analysis instead shows a shared problem across different regimes:[714] AI-supported scoring may improve prediction, but it can also convert alternative data, supervisory risk settings, and managerial risk appetite into adverse outcomes that consumers cannot test.[715] U.S. ECOA and FCRA practice confirms both the value and limits of reason-giving.[716] Specific adverse-action reasons and dispute rights matter, yet factor labels remain incomplete when data sources, correlations, and vendor-controlled logic remain hidden.[717] A structured right to explanation should therefore preserve legitimate risk assessment, bank

708. *See* Citron & Pasquale, *supra* note 24, at 8–9 (explaining that lenders initially relied on retail and banking staff, and later specialized finance companies, to judge borrowers' creditworthiness, before Fair, Isaac & Co. introduced a three-digit numerical credit score in 1956).

709. *See id*. at 1 (explaining that predictive scoring systems determine access to "crucial opportunities. . . including the ability to obtain loans, work, housing, and insurance," rather than functioning as a mere technical convenience).

710. *See id.* at 3, 19 (explaining that "powerful entities. . . turn individuals into ranked and rated objects," and warning that without law and due process in this field, "we are essentially paving the way to a new feudal order of unaccountable reputational intermediaries").

711. *See id.* at 28 (explaining that even where an algorithm's code remains confidential, "it is still possible for independent third parties to review it," so accountability does not depend on full public understanding of how the system works).

712. *See* Wachter et al., *supra* note 356, at 842, 845 (explaining that data controllers have an interest in withholding algorithmic details to protect trade secrets and prevent gaming, and that "the internal state of the algorithm can consist of millions of variables intricately connected in a large web of dependent behaviours").

713. *See supra* notes 185–98 and accompanying text (developing the dual evaluative-personality and intangible-property nature of credit rights, and the debate over recognizing an independent credit right, as the doctrinal basis for treating the right to algorithmic explanation as a hybrid consumer-protection guarantee).

714. *See generally supra* notes 429–47 and accompanying text (comparing the EU, Chinese, and U.S. approaches to algorithmic explanation and identifying a shared problem: notice and access rights that fall short of enabling meaningful contestability).

715. *See supra* notes 122–26 and accompanying text (discussing how managerial risk appetite is transmitted into automated credit models through board-approved limits, approval thresholds, and pricing bands); *see also supra* notes 137, 542 and accompanying text (discussing how alternative data and vendor-controlled scoring practices remain difficult to trace and "practically untested" by affected consumers).

716. *See supra* notes 500–29 and accompanying text (discussing ECOA, Regulation B, FCRA, and CFPB guidance establishing specific reason-giving duties for adverse credit actions); *see also supra* notes 530–38 and accompanying text (discussing the institutional gaps limiting the U.S. reason-giving regime).

717. *Id*.

supervision, capital discipline, and trade-secret protection while requiring adverse credit decisions to rest on specific, usable, and reviewable reasons.[718] This frame clarifies why traditional doctrinal routes underdeliver.[719] A contract can legitimate scoring through assent without affecting the epistemic core;[720] therefore, disputes are decided on authorization rather than contestability.[721] Tort routes then inherit the problem of proof, because the claimant is asked to establish fault and causation while the defendant controls the evidence that can test the inference.[722] The proposed construction answers this by specifying duty bearers, trigger conditions, and procedural steps and by treating explanation as a contestability device rather than the compelled publication of the source code or model weights.[723]

A further implication is that explanation is not inherently emancipatory.[724] If demanded at maximal depth by default, it can collapse into empty formalism through unusable technical artifacts or regulatory paralysis through broad trade secret claims.[725] A legally defensible approach is therefore a content gradient, anchored in proportionality but operationalized through intelligibility and remedy, such that the depth of the explanation increases with the degree of automation and the severity of impact.[726] This also implies that an explanation cannot be treated as a one-off message.[727] Rather, it must be embedded into the product

718. *See supra* notes 112–28 and accompanying text (discussing how bank supervisory frameworks, including EBA loan-origination governance, the EU Capital Requirements Regulation, U.S. model-risk guidance, and Chinese capital-management measures, embed credit-scoring models within risk assessment, capital discipline, and managerial risk-appetite governance); *see also supra* note 80 and accompanying text (explaining that trade-secret constraints and model-level outputs do not by themselves give consumers reviewable reasons for adverse credit outcomes).

719. *See generally supra* notes 222–323 and accompanying text (discussing the contractual and tort based doctrinal routes to challenging automated credit scoring decisions and their evidentiary and interpretive limitations).

720. *See supra* notes 248–62 and accompanying text (explaining that institutions invoke contract terms and business necessity to avoid justifying scoring logic, leaving the reasoning "evidentially insulated" from challenge).

721. *Id*.

722. *See supra* notes 263–71 and accompanying text (explaining that tort claimants must prove "illegality, harm, causation, and fault" while the decisive information sits within an "opaque, defendant-controlled scoring pipeline").

723. *See generally supra* notes 555–600, 634–52, 688–707 and accompanying text (setting out the proposed legal construction of the right to algorithmic explanation, including the right holder and duty-bearer analysis in Part VI.A, the trigger-threshold framework in Part VI.C.2, and the procedural disclosure design in Part VI.D).

724. *See supra* notes 655–59 and accompanying text (explaining that a defensible explanation framework must reject both extremes, "dumping technical artifacts that users cannot parse and offering vague assurances that users cannot dispute").

725. *See supra* notes 655–59, 674–83 and accompanying text (explaining that user-facing disclosure must avoid "dumping technical artifacts that users cannot parse," and noting that "controllers routinely invoke trade secrets as a general justification to resist access requests").

726. *See supra* notes 630, 637–52, 658 and accompanying text (explaining that proportionality is "the more stable normative anchor" for algorithmic explanation, that a defensible gradient rests on "intelligibility, credibility, and remedy," and that the relevant trigger is whether a score "alters the legal position or economic interests" of the affected person).

727. *See supra* notes 684–87 and accompanying text (explaining that "the governance architecture must shift from individual access to institutional inspection," with rights effectiveness "preserved through supervised balancing rather than abandoned through opacity").

architecture and be coupled with repeatable oversight, including audit and supervision mechanisms that do not depend on the resources of a single applicant.[728]

Future work should move in at least three directions. First, empirical legal research should test explanation contents against the functions the law expects them to serve, especially error correction, nondiscrimination scrutiny, and effective appeal.[729] This is not engineering in a legal frame.[730] Rather, it is an assessment of the explanation formats that make contestability usable for credit consumers under real constraints of literacy, time, and power.[731] Counterfactual and feature contribution disclosures represent promising directions because they can be checked against logs and decision pipelines;[732] however, they are legally adequate only if they can reveal the decisive drivers rather than that which is convenient to disclose.[733] Second, scholarship should focus on smaller and more heterogeneous scoring ecosystems, where scoring is embedded in platform governance, thin file lending, renting, employment screening, or buy-now pay-later products.[734] These settings sharpen the multi-actor chain problem and make evasion through nominal human involvement more likely.[735] Third, institutional designs should be developed that reconcile meaningful explanations with legitimate confidentiality and security interests.[736] The priority is to ensure confidentiality that preserves scrutiny through trusted third-party review, court-supervised disclosure with protective measures, and standardized audit logs,

728. *Id.* (explaining the proposed two-track design: a user-facing explanation alongside "a regulator- or court-facing dossier that can be inspected under confidentiality safeguards when the dispute turns to deeper technical or comparative claims").

729. *See supra* note 67 and accompanying text (identifying error correction, discrimination, and contestability as the functions the right to explanation is meant to serve).

730. *See supra* notes 325–26, 331 and accompanying text (explaining that "the decisive question is therefore not whether an algorithm can be fully rendered transparent in engineering terms but rather whether it can be explained to a degree that serves the legal functions of justification, contestation, and oversight," and that "legal explanation is not synonymous with reverse engineering").

731. *See supra* notes 611, 616–17 and accompanying text (explaining that explanation formats must be designed for realistic consumer use, translating complex dependence into a form "a layperson can still use," expressed "in plain language" through "interactive pathways," precisely because "the borrower is expected to act on the explanation" despite limited technical background and capacity).

732. *See supra* notes 612–13 and accompanying text (explaining that a disclosure identifying which factual changes would have altered the score is particularly defensible because it can be checked against the controller's actual pipeline and logs).

733. *See supra* notes 370–72 and accompanying text (explaining that counterfactual disclosures can satisfy a controller's duty while still concealing the actual decisive drivers behind a clear but incomplete narrative).

734. *See* Danielle, *supra* note 34, at 1 (observing that predictive scoring already extends beyond credit into domains such as employment and tenant screening).

735. *See supra* notes 69, 171, 183 and accompanying text (discussing how a reviewer who merely rubber stamps a system output can evade automated-decision protections by exploiting a multi-actor, socio-technical chain).

736. *See supra* notes 684–87 and accompanying text (describing a two-track institutional design, a user-facing explanation paired with a confidential regulator- or court-facing dossier, that reconciles contestability with legitimate secrecy interests).

making deeper verification possible without turning explanation into compelled publication.[737]

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.*

Data availability

No data was used for the research described in the article.

Funding

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. T2600237, the Early Career Scheme (ECS) of the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China, under Project No. 26601126, and the research funding under the HKUST-DXM AI for Finance Joint Laboratory (DXM25EG01).

737. *See supra* notes 426, 613, 686–87 and accompanying text (identifying a neutral expert panel review, court-supervised disclosure under confidentiality safeguards, and log-based inspection mechanisms as the specific verification channels developed earlier).

ARTIFICIAL INTELLIGENCE / LAW

# Center for Artificial Intelligence for Law

人工智能與法律研究中心




PUBLICATION DETAILS

**University of Illinois Journal of Law, Technology & Policy**

2026 · Volume 2026 · Issue 2 · pp. 37–115

RECOMMENDED CITATION

Yidan Huang and Sirui Han*, 'Consumer Protection in Artificial Intelligence–Governed Credit Markets: Algorithmic Bias and the Right to Explanation' (2026) University of Illinois Journal of Law, Technology & Policy, Volume 2026 – Issue 2, pp. 37–115.



* Corresponding author: Sirui Han.


香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

人工智能與法律研究中心
Center for Artificial Intelligence for Law