

Beyond Genetic Discrimination: Emerging Challenges in the Age of Artificial Intelligence

Li Zhenkang

Abstract: With the deep integration of artificial intelligence and biotechnology, the paradigm of genetic discrimination is undergoing a fundamental transformation. Rather than relying on specific genetic markers, AI systems may infer individuals' underlying biological risks through proxy data. This emerging form of discrimination, conceptualized in this article as “Algorithmic Proxy Genetic Discrimination” (APGD), is predictive, opaque, and structural. However, China’s existing Personal Information Protection Law and fragmented antidiscrimination framework remain insufficient to address this problem. In addition, APGD raises profound ethical and social challenges. To mitigate these risks, this article proposes a tiered regulatory approach: (1) controlling the misuse of proxy variables at the data input stage; (2) strengthening transparency and auditing in algorithmic decision-making; and (3) establishing an outcome-oriented mechanism for anti-discrimination liability and rights protection.

Keywords: APGD; Proxy Data; Anti-Discrimination Law; Bioethics; AI

Funding: The author received no financial support for the research, authorship, and publication of this article.

Competing Interests: The author declares no potential conflicts of interest.

Author Affiliations: Li Zhenkang is a PhD candidate in Medical Law at the School of Civil and Commercial Law, Southwest University of Political Science and Law, one of China’s leading institutions for legal education. He is currently a visiting doctoral scholar at the University of Illinois Urbana-Champaign College of Law.

Introduction

Genetic discrimination is a distinctive subtype of discrimination that has emerged primarily alongside advances in commercial DTC genetic testing and gene-editing technologies.¹ This issue has become increasingly salient as DTC genetic testing has moved from a niche biomedical service to a rapidly expanding consumer market.² According to industry reports, the global DTC genetic testing market was valued at approximately USD 2.17 billion in 2025 and is projected to reach approximately USD 11.02 billion by 2035.³ In fact, genetic discrimination is not a new phenomenon. The first case of genetic discrimination emerged in the late 1970s in the United States, particularly in private insurance and employment.⁴ Since then, many countries across Europe, Asia, and Oceania have successively enacted anti-discrimination laws to prevent such practices.⁵ For example, a civil service recruitment dispute in Foshan, Guangdong, in 2010, in which a candidate was disqualified for carrying the thalassemia gene and later filed an administrative lawsuit, is widely regarded as China's first genetic discrimination case.⁶ Following intense public and medical debate, the local government revised the physical examination standards in 2011, removing genetic testing for the thalassemia trait.⁷ This historical trajectory underscores a critical point: traditional genetic discrimination was relatively visible, and its direct association with physical biomarkers made the resulting injustice easier to identify and address through legal and policy reform.

However, by the late 1990s, legal scholars had already begun to warn that the more serious threat in the future would lie beyond these visible forms of discrimination.⁸ For example, Susan M. Wolf argued that restricting access to physical DNA alone would not prevent “geneticism.”⁹ Today, that warning has materialized through a technological pathway she could scarcely have foreseen. With the deep integration of AI and biotechnology, the paradigm of genetic discrimination has been fundamentally transformed.¹⁰ Rather than relying solely on visible physiological differences or specific genetic markers, AI systems combine electronic health records (EHR), behavioral data, and other proxy indicators to infer probabilistic risks relating to future illness, incapacity, or cost.¹¹ In this process, individuals are no longer assessed primarily based on their present ability or actual conduct; they are sorted according to predicted biological risk.¹²

To better understand this shift, consider an everyday scenario: During pre-employment screening, a healthy engineer might be flagged as “high risk” by an AI HR system based on his EHR or regional health data. The resulting disadvantage stems not from a manifested disease or identifiable genetic defect, but from a purely algorithmic inference regarding his future biological vulnerability.¹³ More importantly, the scenario has moved beyond the realm of theory; it is an emerging reality in the era of AI. As Anya Prince and Daniel Schwarcz have shown, even when the law prohibits algorithms from using protected traits such as race, gender, or genetic data, they may still maximize predictive power through “rational” proxy discrimination by relying on facially neutral variables correlated with those traits, regardless of discriminatory intent.¹⁴ In other words, genetic discrimination has not disappeared; rather, it is increasingly being reconstituted through AI-driven, big data-enabled forms of predictive screening — a practice I call “algorithmic proxy genetic discrimination” (APGD).

Current legal and theoretical scholarship has yet to adequately respond to this emerging threat. Substantial research has addressed traditional genetic discrimination, general algorithmic discrimination, and algorithmic proxy discrimination (APD), and legislative protections for genetic data, such as the *Genetic Information Nondiscrimination Act of 2008* (GINA) in the United States, are already in place.¹⁵ However, existing scholarship has largely failed to address the intersection between algorithmic proxy mechanisms and genetic discrimination, leaving a critical regulatory gap. Properly identifying and resolving this new issue is crucial for the effective future regulation of this novel form of discrimination.¹⁶ Although this Article draws upon the antidiscrimination frameworks of China and the United States as its primary analytical backdrop, the institutional challenges posed by APGD are inherently universal.¹⁷ As such, the normative proposals and legal analysis developed herein offer highly relevant, forward-looking insights for global policymakers.

Methodology

This article adopts a doctrinal and conceptual legal analysis to frame “Algorithmic Proxy Genetic Discrimination” (APGD) as a distinct form of discrimination, rather than as a mere extension of traditional Genetic Discrimination (GD) or a simple subcategory of Algorithmic Proxy Discrimination (APD). It deliberately draws on literature on genetic discrimination, algorithmic fairness, and health data governance, selected because these bodies of scholarship most directly address the intersection between genetic risk, proxy inference, and data-driven decision-making. Empirical examples, such as the Obermeyer algorithm study, are used not as comprehensive empirical evidence, but as illustrative cases showing how proxy variables may reproduce forms of biological exclusion without relying on direct genetic testing. The legal analysis focuses on GINA, *China’s Personal Information Protection Law* (PIPL), and other selected policy materials because these instruments represent major comparative regulatory approaches to genetic information, health data, and algorithmic governance.

Contributions and Structure of the Article

This Article makes three main contributions. First, it identifies the regulatory gap created when AI systems infer genetic or biological risk through proxy variables without directly using genetic data. Second, it offers a conceptual innovation by defining “algorithmic proxy genetic discrimination” (APGD) as a distinct form of discrimination that bridges, yet differs from, traditional genetic discrimination (GD) and algorithmic proxy discrimination (APD). Third, it proposes a tiered doctrinal response that combines proxy-data control, algorithmic transparency and auditing, and outcome-based anti-discrimination liability.

In this regard, this article proceeds as follows: Part II examines the paradigm shift from traditional genetic discrimination to algorithmic proxy genetic discrimination. It clarifies why APGD constitutes a distinct new category of discrimination, identifies its core characteristics, and examines the key driving forces behind its emergence. Part III analyzes the emerging challenges posed by APGD in the Chinese context, focusing on its ethical risks, legal loopholes, and broader social implications for multi-stakeholder governance and the balance between governance and innovation. Part IV proposes a tiered regulatory approach for addressing APGD. Finally, Part V concludes.

The Paradigm Shift from Genetic Discrimination to Algorithmic Proxy Genetic Discrimination

APGD as a Distinct Form of Discrimination in the AI Era: Why It Is Not Merely a Subset of GD or APD.

In this article, APGD is defined as a novel form of discrimination in which AI systems, without relying on physical genetic samples or formal genetic test results, mine and analyze an individual's proxy data to infer underlying genetic traits or biological risks, thereby producing outcomes that are tantamount to genetic discrimination in effect. These algorithmic inferences are then used to subject individuals to unfair treatment in the allocation of essential socioeconomic opportunities, such as employment, insurance, and healthcare.¹⁸

It is necessary to further clarify why APGD should be treated as an independent analytical category rather than subsumed under existing frameworks. First, APGD is doctrinally distinct from traditional Genetic Discrimination. Current legal regimes addressing genetic discrimination, with GINA being the most prominent example, generally confine protection to explicitly defined genetic information, such as DNA test results.¹⁹

In other words, under 42 U.S.C. § 2000ff of GINA, traditional genetic discrimination is generally premised on the acquisition of substantive biological genetic data, such as DNA sequencing or chromosomal analysis.²⁰ By contrast, the underlying logic of algorithmic proxy genetic discrimination (APGD) departs from this reliance on physical DNA, instead employing predictive algorithms to infer health risks from non-genetic proxy data. Therefore, if legal analysis rigidly adheres to the traditional GD framework, using GINA as an example, will the law inevitably be consigned to an institutional blind spot marked by regulatory failure and statutory silence? More specifically, should the definition of “genetic information” under GINA be reconsidered or modestly expanded to cover proxy data that can be used to infer genetic traits or biological risks?

Second, APGD cannot be reduced to a subset of algorithmic discrimination and algorithmic proxy discrimination (APD). Algorithmic discrimination is the broader category under which APD falls;²¹ APD is a specific mechanism within algorithmic discrimination, whereby algorithms reproduce discrimination based on protected social attributes through the use of proxy variables. More specifically, general algorithmic discrimination usually concerns how automated systems reproduce social bias, such as unequal treatment based on race, gender, age, disability, or socioeconomic status.²² At the level of algorithmic operation, APD focuses more specifically on how algorithms indirectly reproduce such bias through facially neutral proxy data.

Although the framework of APD recognizes the use of neutral proxy data to reproduce bias, it remains primarily concerned with discrimination based on sociological attributes, such as race, gender, or socioeconomic status.²³ However, it does not adequately account for a distinct form of harm arising from predictions about an individual's future biological condition. Unlike APD, which focuses on relatively static sociological attributes such as race and gender, APGD seeks to

infer human biological vulnerability and the future probability of disease onset, treating genetic risk as a dynamic and probabilistic attribute of life and health.

In addition, there is a critical divergence in the underlying ethical harms (Social Bias vs. Geneticism). APD tends to amplify entrenched social prejudices, such as the marginalization of vulnerable groups. APGD, however, triggers a far more profound bioethical crisis: geneticism and biological determinism. Beyond merely depriving individuals of employment opportunities, APGD fundamentally subjects individuals to social stratification based on their perceived “biological destiny”.²⁴ This practice deeply implicates medical ethics and human dignity, thereby transcending the conventional boundaries of algorithmic fairness discourse.²⁵ To better clarify the paradigm shift from GD to APGD, Table 1 compares their key differences in data sources, mechanisms, and legal implications.

To sum up, although APGD shares certain features with traditional genetic discrimination and APD, it cannot be fully subsumed under either framework. Unlike traditional genetic discrimination, APGD does not depend on explicit genetic data; unlike APD, it concerns inferred biological vulnerabilities and prospective health risks rather than static sociological attributes. Therefore, APGD is a new form of discrimination in the AI era. Failure to recognize APGD as a distinct category leaves a profound form of discrimination unprotected by existing laws.

A Summary of the Core Characteristics of APGD

It is crucial to emphasize that APGD is a new form of discrimination in the AI era. It represents a hybrid and distorted evolution of both genetic discrimination and APD. APGD has three main core characteristics:

(1) It is proxy-based. Even where the decision-making process does not directly collect or use genetic information, AI systems may still infer an individual’s underlying biological susceptibility from proxy data.^{26,27} Legally, this exposes a structural limitation. Because current regulations mainly focus on clearly defined sensitive data, such as genetic information. However, AI systems can rely on proxy data to infer similar risks, allowing discriminatory practices to bypass existing legal controls.

(2) It is AI-driven. Once proxy data is collected, it is processed through AI systems as mediating mechanisms that generate probabilistic assessments of biological vulnerability and convert risk signals into preemptive adverse outcomes.²⁸ Due to the inherent “black box” nature of AI systems, their decision-making processes are often insufficiently transparent, making it difficult for affected individuals to understand why they have been classified as high-risk, let alone to meet the evidentiary burden required to legally challenge such determinations.²⁹

(3) It is structural. This new discrimination rarely operates in isolation; instead, it readily interacts with preexisting disadvantages associated with poverty, illness, disability, or geography, and may reinforce and entrench social inequality through continuous data-driven feedback loops.³⁰ It should be noted that APGD is still primarily applied in the fields of employment, insurance, and healthcare.

Table 1: The Paradigm Shift from Traditional Genetic Discrimination to APGD

Comparison	GD	APGD
Data Source	Physical biological samples, direct DNA sequencing	EHR, digital behavioral trajectories, consumer data, zip codes, online searches.
Mechanism	Exclusion based on clear biological facts or diagnosed genetic traits.	Complex mathematical probability inference and high-dimensional feature matching extracting proxy variables via machine learning (ML).
Transparency	High. Victims generally know they face exclusion due to submitting genetic or medical reports.	Extremely low. Due to black-box algorithms and mathwashing, causal links are nearly impossible to prove.
Legal Regulation	Partially regulated. Covered by specific laws (e.g., GINA) or strict medical ethics guidelines.	Legal vacuum. Current privacy and antidiscrimination frameworks struggle to address data-correlated predictive penal

The Driving Forces Behind the Shift Toward APGD

Capital's Strategic Exploitation of the Gray Areas in Anti-Discrimination Law

Laws have a strong deterrent effect.³¹ With the enactment of anti-discrimination laws exemplified by the U.S. GINA, employers and health insurers were expressly prohibited from directly obtaining and using individuals' DNA sequences in hiring or underwriting decisions.³² This strict regulation of "substantive biological data" significantly increased the legal risks and compliance costs associated with traditional genetic discrimination. In other words, the enactment of GINA reduced the risk of genetic discrimination to some extent and, in turn, helped alleviate public concerns in the United States about the misuse of genetic information. One empirical finding shows that prior to GINA's enactment, approximately 26% of high-risk individuals refused to undergo genetic testing due to concerns about insurance denial or employer discrimination; by 2013, five years after its implementation, this proportion had fallen to about 3.2%, indicating a significant downward trend.³³

However, the underlying drive of capital to minimize risk and maximize profit did not disappear.³⁴ Given that APGD is predominantly deployed in commercial sectors such as employment, healthcare, and insurance, it is inextricably linked to the concept of commercial capital. Unless the inherent profit-driven logic of "minimizing risk and maximizing profit" changes, laws will only force capital's discriminatory practices to mutate through technological upgrades. Consequently, these practices may cloak themselves in the guise of big data, becoming

increasingly insidious and slipping through existing legal loopholes. A clear example is GINA: although it regulates the acquisition and use of explicitly defined “genetic information,” its statutory definition does not clearly extend to proxy data that can be used by AI systems to infer genetic traits or biological risks.³⁵ This displacement effect became the fundamental driver of technological evolution: Discrimination was not eliminated; rather, it was transformed into forms capable of evading existing regulation.

Mathematical Cleansing: The Illusion of Objectivity

APGD operates on the principle of “mathwashing”—a term popularized by data scientist Fred Benenson to describe the false assumption that mathematically grounded algorithms are inherently objective.³⁶ In fact, because ML or AI models optimize solely for predictive accuracy without inherent moral constraints, they easily absorb the human biases and historical inequalities embedded within their training data. A stark example is Amazon’s scrapped AI recruiting tool: trained on 10 years of male-dominated tech industry resumes, the algorithm mathematically concluded that male candidates were preferable and automatically downgraded resumes containing the word “women”.³⁷ This perfectly illustrates how AI can blindly automate historical discrimination. Therefore, this allows decision-making institutions to easily pass the buck to algorithms. They can dodge the accusation of discrimination by calling it data-driven business intelligence based on neutral data. This opacity leaves victims with little way of knowing whether they were denied services because an algorithm had effectively classified them as resembling a group associated with certain genetic defects.³⁸

Empirical Mapping: Proxy Data Entrenches Inequality Again

The danger of proxy variables in medical and social sorting is vividly demonstrated by empirical studies. In a landmark 2019 study published in *Science*, researchers led by Ziad Obermeyer examined a commercial algorithm used to manage healthcare for millions of patients.³⁹⁻⁴⁰ The algorithm exhibited severe racial bias because it used “historical healthcare costs” as a convenient proxy for “future health needs”. Due to systemic inequalities and unequal access to care, Black patients historically incurred lower healthcare costs than White patients with the same level of actual illness. However, the algorithm falsely inferred that Black patients were healthier, cutting by more than half the number of Black patients identified as eligible for additional clinical support. This argument is not intended to claim that all proxy-based prediction produces APGD. Rather, it shows how proxy-based inference may create APGD-like risks when used in exclusionary decision-making without transparency, justification, or safeguards.

If this logic of algorithm proxy discrimination (APD) is mapped onto the framework of APGD, the implications become deeply troubling. Just as “healthcare costs” functioned as a flawed and biased proxy for “health needs,” algorithms can readily use variables such as gym attendance, dietary purchases, or zip codes as proxies for genetic vulnerability. In this way, APGD can quietly reproduce and entrench structural inequality, giving rise to a form of genetic redlining in the digital age.⁴¹

The Emerging Challenges Posed by APGD in the Chinese Social Context

When evaluating the impact of Algorithmic Proxy Genetic Discrimination (APGD), one must recognize that emerging technologies often trigger ethical crises before they generate legal responses. It is precisely in the effort to address these ethical crises that the vulnerabilities of existing legal frameworks and the core challenge of governance become exposed.

The Core Ethical Challenges of AI-Driven “New Discrimination”

The Erosion of Fairness and Equality

The erosion of fairness is most clearly reflected in the shift in allocation standards from equal evaluation based on actual ability to predictive screening based on future risk.⁴² Under a traditional conception of fairness, individuals should be assessed primarily on the basis of their present physical condition, actual capacity, or real conduct. However, APGD imposes a form of “algorithmic determinism.” Once proxy data is incorporated into ML or AI, individuals may be identified, categorized, and excluded in advance—even though no actual harm has occurred and no present incapacity can be shown. Fairness thus no longer consists in treating real individuals equally; instead, it is displaced by a technical logic of preventing uncertain future risks.⁴³

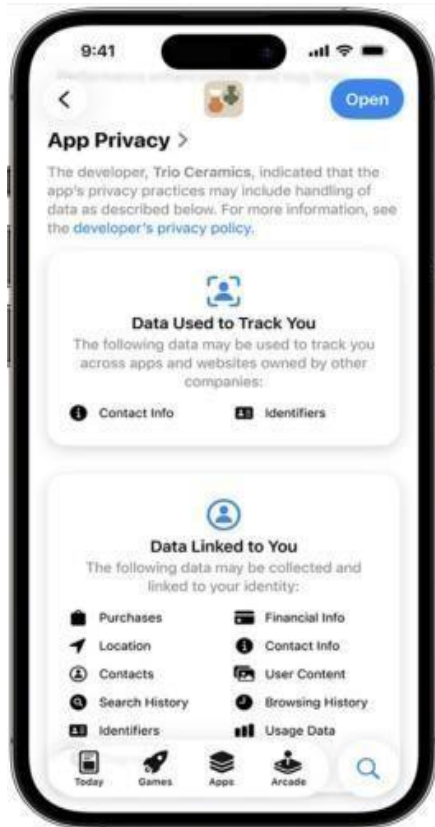
More importantly, during the training process, algorithmic models often absorb the biases and inequalities embedded in existing social data and reproduce stereotypical representations of certain groups as “high risk.”⁴⁴ For example, if the training data come from an industry long dominated by men, the algorithm may learn the implicit bias that male candidates are better suited for technical positions. Even without directly using gender as a variable, the system may still rely on indirect signals—such as work experience, educational background, or patterns of word choice—to classify female applicants as a poorer fit. So, once members of particular families or regional populations are repeatedly identified as groups likely to generate high healthcare costs, these probabilistic judgments may gradually harden into institutionalized labels.⁴⁵ Over time, algorithms may not only magnify existing social prejudice, but also further entrench structural inequality through feedback loops, leaving disadvantaged groups persistently marginalized.

The Undermining of Informed Consent and Individual Autonomy

In the course of algorithmic decision-making, AI systems may not rely on DNA data itself, but instead infer an individual’s future health risks by combining proxy data.⁴⁶ Considering a frequently used app interface, as shown in Figure 1, users of shopping apps often consent to the collection of their “location data” and “shopping habits” in order to access delivery services or personalized product recommendations. In their perception, this merely involves exchanging relatively harmless data for convenience. However, within the opaque “black box” of algorithmic systems, AI conducts deep correlation and integrative analysis of these fragmented data points.⁴⁷ By aggregating records such as late-night purchases of high-sugar fast food and over-the-counter medications, together with behavioral patterns indicating a lack of outdoor activity, the system can, even without directly accessing blood samples or family medical history, generate highly accurate predictions regarding an individual’s future risk of cardiovascular diseases or diabetes.⁴⁸

The core problem is that, even when individuals consent to the collection of certain types of data, such as location information or shopping habits, they often do not truly understand how those data may be combined and what they are used for. In this sense, what appears to be “consent” on the surface does not amount to real informed consent.

Figure 1: Example of APP-Associated Data⁴⁹



Predatory Marketing and Exploitation of Vulnerability

While traditional precision marketing seeks to predict what consumers want, marketing driven by genetic inference directly targets what they fear. When algorithms leverage proxy data—such as purchasing habits, search histories, and geolocation—to infer an individual’s latent genetic traits or health risks, the fundamental nature of precision marketing undergoes a radical transformation.⁵⁰ It ceases to be mere “product recommendation” and mutates into the targeted predation of human physiological and psychological vulnerabilities.

Should an algorithm deduce that a consumer has a heightened risk for Alzheimer’s disease or a specific hereditary cancer, the marketing system may systematically bombard them with unproven, exorbitant health supplements, aggressive preventive medical interventions, or overpriced private health insurance. Trapped in a state of severe information asymmetry, consumers remain oblivious as to why they are targeted by such advertisements. Consequently, their health-related anxieties are subconsciously amplified, driving irrational consumption

decisions. From an ethical standpoint, this constitutes the covert exploitation of vulnerable populations—specifically, those rendered psychologically fragile by their potential disease risks.⁵¹

Challenges to the Current Legal Landscape in China

PIPL: Struggling to Keep Pace with AI-Driven Genetic Inference

As discussed above, APGD gives rise to the effects of genetic discrimination. But in terms of anti-discrimination laws, China currently faces a legislative vacuum. Compared to the limited jurisdictional scope of the U.S. GINA, there remains a significant legislative lacuna in China's legal system concerning dedicated regulations for bioethical discrimination. Why has China yet to enact specific, direct legislation addressing bioethical discrimination? (1) The *Personal Information Protection Law of the People's Republic of China (PIPL)* categorizes genetic information as “sensitive personal information,” mandating separate consent and the execution of a Personal Information Protection Impact Assessment.⁵² (2) The primary impetus behind the enactment of GINA in the U.S. stemmed from its highly commercialized health insurance market, where the public harbored profound concerns that genetic predispositions might lead to the loss of commercial coverage.⁵³ In contrast, China implements a broad-based basic medical insurance system; an individual's genetic information does not compromise their access to fundamental public healthcare.

However, unlike the United States, China, much like the EU, tends to address these issues primarily through a data governance framework rather than an anti-discrimination framework.⁵⁴ In the context of mainland China's civil law system, genetic data falls within the regulatory framework of the *Personal Information Protection Law of the People's Republic of China (PIPL)*. The PIPL, which came into force on November 1, 2021, is a comprehensive statute enacted to protect personal information rights and interests and to regulate personal information processing activities. Article 28 of the PIPL classifies information such as “biometric identification” and “medical health” information as sensitive personal information, and defines sensitive personal information as personal information that, once leaked or unlawfully used, is likely to result in harm to the dignity of natural persons or endanger their personal or property safety.⁵⁵ Although Article 28 does not expressly list “genetic data” as an independent category, genetic data usually possesses biometric, medical health-related, and identity-identifying attributes. Therefore, as a matter of legal interpretation, it should be included within the protective scope of sensitive personal information.

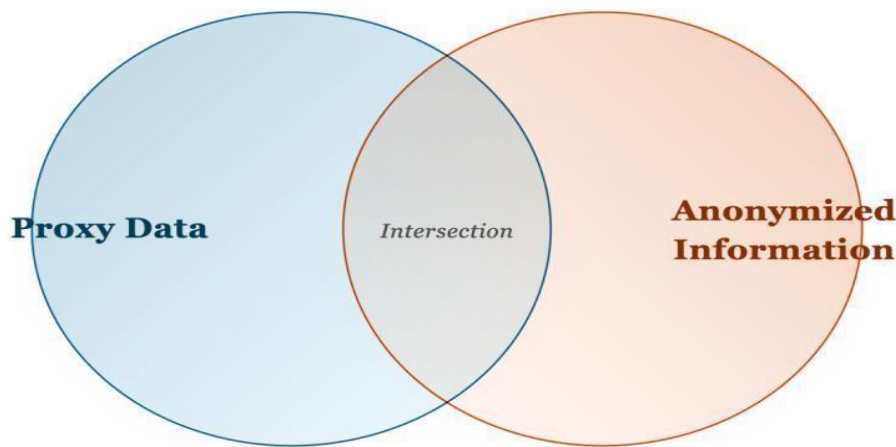
By the same logic, should non-genetic proxy data involved in APGD also be brought within the regulatory scope of the PIPL? If not, would such data fall outside the Personal Information Protection Law and therefore remain beyond its effective regulatory reach?

On this basis, I identify a deficiency in the current framework concerning the legal consequences of anonymization. This concerns the legal consequences of anonymization. Under Article 73 of the Personal Information Protection Law, anonymization requires that personal information be processed in such a way that a specific natural person can no longer be identified and the information cannot be restored.⁵⁶ Once data satisfies this standard of full anonymization, it is, in principle, removed from the scope of the PIPL. Yet this exclusion may create a regulatory

loophole. In China, when proxy data used for APGD is exploited by commercial actors, it will almost inevitably be processed in anonymized form. This raises a further question: what is the relationship between APGD proxy data and anonymized data?

For a clearer illustration, Figure 2 shows the relationship between the two. Anonymized pharmaceutical purchase records provide a useful example of their overlap. In other words, this overlapping category may formally evade the regulatory logic of the PIPL's anonymization rules, which are intended to create space for legitimate commercial innovation. However, despite the loss of direct identifiability, such data may still enable AI systems to infer biological vulnerabilities or future health risks, thereby generating the very discriminatory risks that APGD seeks to capture.

Figure 2: The Set Relationship Between Anonymized Data and Proxy Data.
(Source: Author's own elaboration.)



Fragmented Anti-Discrimination Rules: Inadequate Protection Against APGD

As noted above, China does not have anti-discrimination legislation comparable to the U.S. GINA. Instead, anti-discrimination in China is mainly reflected in a fragmented set of constitutional principles, labor law provisions, PIPL rules, and sector-specific regulatory norms.

However, traditional anti-discrimination rules generally rest on two assumptions: first, that the prohibited ground of discrimination can be identified; and second, that the unfair treatment is sufficiently visible to be proven. APGD challenges both assumptions.⁵⁷ It operates through prediction, proxy variables, and opaque algorithmic processes, making it difficult to determine

precisely why a person has been disadvantaged or to establish the causal link required under traditional legal frameworks.

Challenges to Social Governance

APGD as a Cross-Sectoral Governance Problem

The new form of discrimination generated by APGD cannot be effectively governed through any single area of law. This is because APGD is not merely a privacy issue or a discrimination issue. Rather, it is a complex governance problem involving law, medicine, artificial intelligence, data governance, and public policy.⁵⁸

From a privacy-law perspective, APGD involves the collection, processing, and secondary use of health-related, behavioral, consumer, and lifestyle data. Yet privacy law alone is insufficient, because APGD often operates through proxy data rather than explicit genetic information. Such data may appear neutral or non-sensitive on its face, but it can still be used to infer biological vulnerability, disease predisposition, or future health risks.

From an anti-discrimination and AI-governance perspective, APGD also challenges existing legal frameworks. Traditional anti-discrimination rules usually depend on the identification of a protected ground and a visible act of unequal treatment, while AI governance focuses on transparency, explainability, accountability, and contestability. APGD complicates both approaches because discriminatory effects may be produced through opaque algorithmic correlations, probabilistic predictions, and seemingly neutral variables. Therefore, APGD should be understood as a cross-sectoral governance challenge requiring coordination among personal information protection, anti-discrimination law, medical ethics, AI governance, consumer protection, insurance regulation, employment law, and public health policy.⁵⁹

Counterarguments: The Tension Between Governance and Innovation

One possible objection to this article's argument is that predictive risk assessment should not be regarded as inherently discriminatory. In fields such as healthcare, insurance, employment, and public administration, predictive models may serve legitimate functions. For example, risk prediction may help identify high-risk patients, optimize the allocation of medical resources, and support preventive medicine.⁶⁰ Actuarial assessment in insurance has long been used to classify risk pools and maintain the sustainability of insurance systems. Predictive tools in public administration may also be used to detect fraud, anticipate public health risks, or improve the distribution of welfare resources.⁶¹ Therefore, the use of proxy variables and predictive analytics is not necessarily objectionable in itself.

This objection is important because a complete prohibition on predictive risk assessment would be neither realistic nor desirable. In the age of artificial intelligence, predictive analytics has become an important driver of technological progress, institutional innovation, and business

development. If the law were to adopt a blanket prohibition, it might unduly suppress beneficial innovation, weaken incentives for enterprises to develop socially valuable AI applications, and reduce institutional capacity to respond to complex risks. Accordingly, the problem addressed by APGD should not be understood as a general rejection of prediction, risk assessment, or AI innovation.⁶² Rather, it concerns how to strike a balance between encouraging technological progress and preventing biological-risk inferences from being used for exclusionary or discriminatory treatment.

It is precisely against this background that China has advocated “agile governance” in the development of AI.⁶³ Agile governance itself is a common concept in international technology governance, digital governance, and AI governance. It emphasizes the use of flexible, dynamic, and adjustable regulatory approaches to respond to rapidly changing technological risks. China has incorporated this concept into its own AI governance framework and has recognized it as one of the principles of AI governance.⁶⁴ Why, then, has China proposed or adopted agile governance? There are three main reasons. First, AI technology develops so rapidly that unified legislation may easily lag behind technological change. Second, China needs to balance development with security in AI governance. Third, AI-related risks are highly context-specific.

The real legal problem arises when predictive assessment shifts from supportive intervention to exclusionary classification. The decisive issue is not whether prediction is used, but how it is used, for what purpose, and with what consequences. When inferred genetic or biological vulnerability is used to deny opportunities, impose additional burdens, or inflict other forms of unfair treatment without justification, proportionality, and substantive safeguards, heightened legal scrutiny should be triggered. By contrast, when prediction is used to provide support, improve prevention, or allocate assistance, it may be legally and ethically justified. The purpose of regulating APGD is therefore not to impede innovation, but to ensure that predictive technologies remain compatible with equality, dignity, and the individual’s right to an open future.

Constructing a Well-Calibrated Regulatory Approach for APGD

Controlling the Misuse of Proxy Variables at the Data Input Stage

Although data such as EHRs, physical examination results, place of residence, occupation, and consumption patterns may not appear to constitute genetic data on their face, they can readily function as proxy variables in big-data and machine-learning systems for predicting an individual’s future risk of illness or biological vulnerability. Therefore, addressing APGD cannot be limited to asking whether an algorithm directly uses genetic test results. The central concern is that APGD may operate through algorithmic systems that infer an individual’s genetic risk without ever accessing genetic data itself.⁶⁵

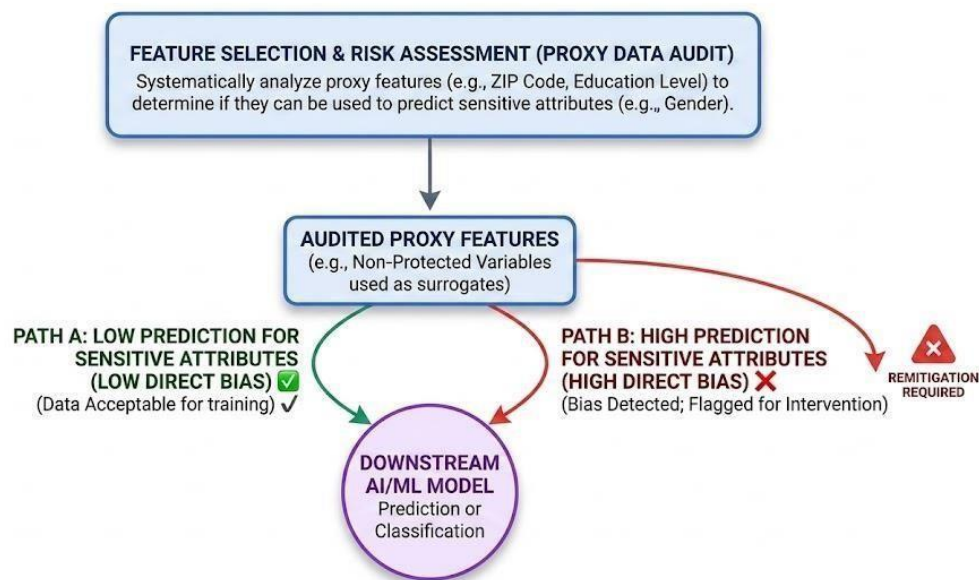
This article argues that a “risk test” should be introduced at the upstream stage of proxy data processing. Even where certain data appear neutral and objective on their face, they should be

subject to regulatory scrutiny if their use in machine-learning or algorithmic systems may produce harmful or discriminatory outcomes.⁶⁶ As shown in Figure 3, the central concern is not merely the nature of the data itself, but the risks generated by its computational use.

It should nevertheless be emphasized that not all proxy data should be subject to blanket restrictions. An overly broad approach would risk overregulation, inhibit socially beneficial innovation, increase compliance costs, and undermine legal certainty. Accordingly, in high-risk sectors such as employment, insurance, and healthcare, the use of high-risk proxy variables should be subject to targeted restrictions, so as to prevent algorithms from engaging in genetic discrimination under the guise of “neutral” data.⁶⁷

Figure 3: Proxy Data Risk Test in AI and ML Systems

(Source: Author’s own elaboration.)



Strengthening Transparency and Auditing in Algorithmic Decision-Making

The hidden nature of APGD lies in the fact that it does not always appear in the form of direct discrimination. Rather, it is often embedded in black-box models, statistical inferences, and seemingly neutral technical processes, through which adverse outcomes are packaged as objective computational results.⁶⁸

In this circumstance, a rational way to regulate this problem is to make sure that harmful decisions cannot be justified merely by saying that “the algorithm produced the result.” If an algorithm is used to deny a person an opportunity, service, or benefit, the decision-maker should be required to explain the main reasons behind the outcome, notify the individual that algorithmic tools were involved, and provide a meaningful opportunity to challenge the result and request human review. At the same time, regulation should not focus only on whether the input data were collected lawfully; it should also examine whether the final outcome is unfair or discriminatory in practice. In high-risk settings such as healthcare, insurance, and employment, algorithmic systems

should therefore be subject to record-keeping, risk assessment, and regular auditing, so that “objective computation” cannot become a shield for hidden discrimination.⁶⁹

An Outcome-Oriented Mechanism for Anti-Discrimination Liability and Rights Protection

Regulating APGD requires more than examining whether data collection is lawful or whether algorithmic procedures are formally compliant. It also requires asking whether a decision, in substance, imposes adverse consequences on an individual.⁷⁰ This argument has two layers. First, regulation cannot stop at the level of formal legality. Even if data collection, data processing, and algorithmic operation formally comply with existing legal requirements, this does not necessarily mean that the resulting decision is substantively fair. Second, the central issue is whether the outcome effectively excludes, disadvantages, or restricts individuals on the basis of inferred biological vulnerability.

For this reason, individuals harmed by algorithmic decisions should not be left to passively accept the result. They should be granted a comprehensive set of procedural and remedial rights, including the right to be informed, the right to object, the right to explanation, the right to correction, and the right to an effective remedy. At the same time, the law should reduce the practical difficulty of proving algorithmic discrimination.⁷¹ One of the central problems in this area is that ordinary individuals cannot see inside the algorithmic black box. Therefore, the burden of proof should not rest entirely on the affected individual. Instead, the law should ease this burden through appropriate evidentiary and procedural mechanisms.⁷²

In Chinese tort law, this remedial framework may be supported by the doctrine of preventive liability. Article 1167 of the Civil Code provides that where a tortious act endangers another person’s personal or property safety, the affected person has the right to request the tortfeasor to assume tort liability, including cessation of infringement, removal of obstruction, and elimination of danger. This provision provides a useful doctrinal basis for addressing APGD.⁷³ Where algorithmic systems create a substantial risk of discriminatory exclusion based on inferred biological vulnerability, the legal response should not wait until concrete damage has already occurred. Instead, preventive remedies should be available to stop, correct, or constrain high-risk algorithmic practices before they produce irreversible harm.

Conclusion

This article proposes a paradigmatic shift from traditional genetic discrimination to Algorithmic Proxy Genetic Discrimination (APGD). It argues that, in the AI era, discrimination may no longer depend on the direct use of genetic test results, family medical history, or explicit genetic data. Instead, algorithmic systems may rely on apparently neutral proxy variables—such as health records, consumer behavior, lifestyle data, or pharmaceutical purchasing patterns—to infer an individual’s future biological risks, thereby producing outcomes that are tantamount to genetic discrimination in effect. APGD therefore represents a new form of discrimination in the age of AI. However, it should not be understood merely as a subset of traditional genetic discrimination or algorithmic proxy discrimination. Rather, APGD occupies an intermediate and

distinctive position: it inherits the normative harm of genetic discrimination, but operates through the technical mechanism of proxy-based algorithmic inference.

The analysis further shows that the emergence of APGD is not merely a product of artificial intelligence, but is also driven by deeper structural factors, including the logic of capital-driven profit seeking and the mechanism of “mathematical cleansing.” As an emerging phenomenon, APGD inevitably gives rise to a series of new challenges. At the legal level, it exposes regulatory gaps within China’s existing normative framework, because current rules on anti-discrimination law, privacy protection, and algorithmic governance cannot fully capture discrimination based on inferred biological risk. At the social level, APGD also raises the difficult question of how to construct multi-stakeholder governance while maintaining an appropriate balance between governance and innovation.

In response, this article proposes a well-calibrated regulatory framework. First, at the data input stage, the misuse of proxy variables should be controlled through a risk-based classification model. Second, during the algorithmic decision-making process, transparency and auditing should be strengthened to prevent apparently neutral models from producing discriminatory outcomes. Third, at the remedial stage, an outcome-oriented mechanism for anti-discrimination liability and rights protection should be established. This mechanism focuses not only on whether data collection and algorithmic procedures are formally lawful, but also on whether the resulting decision imposes adverse consequences on individuals on the basis of inferred biological vulnerability. In Chinese law, this approach may also be supported by Article 1167 of the Civil Code, which provides a doctrinal basis for preventive tort liability where a tortious act endangers another person’s personal or property safety.

Acknowledgements

This article was first drafted during the Spring Festival in Nanking and completed in early March in Chongqing. I am sincerely grateful to Professor Sara Gerke of the University of Illinois Urbana-Champaign College of Law for her helpful comments and guidance. I also thank the anonymous reviewers for their careful and constructive comments, as well as the editorial team for their professionalism and support throughout the review process. I fully recognize that the completion of an article is never solely the result of one author’s individual effort. Rather, it is a process shaped by academic guidance, classroom inspiration, peer review, and editorial support. All errors are my own.

References

- 1 Levin N. A defense of genetic discrimination. *Hastings Center Report*. 2013 Jul;43(4):33-42.
- 2 Horton R, Crawford G, Freeman L, Fenwick A, Wright CF, Lucassen A. Direct-to-consumer genetic testing. *Bmj*. 2019 Oct 16;367.; Majumder MA, Guerrini CJ, McGuire AL. Direct-to-consumer genetic testing: value and risk. *Annual Review of Medicine*. 2021 Jan 27;72(1):151-66.
- 3 Precedence Research. Direct-to-consumer genetic testing market size, share, and trends 2026 to 2035 [Internet]. Ottawa: Precedence Research; 2026 Jan 19 [cited 2026 Mar 30]. Available from: <https://www.precedenceresearch.com/direct-to-consumer-genetic-testing-market>
- 4 Joly Y, Dalpe G. Genetic discrimination still casts a large shadow in 2022. *European Journal of Human Genetics*. 2022 Dec;30(12):1320-2. 5 *Id*.
- 6 Yang X. Genetic Rights and Their Jurisprudential Foundations: A Study of China’s First Genetic Discrimination Case and the “Gene-edited Babies” Incident. *Journal of Humanities, Arts and Social Science*. 2025 Jul 8;9(6).

- 7 Lin J. China's first genetic discrimination case explores social fairness [Internet]. China Youth Daily. 2010 Jan 13 [cited 2025 Dec 12]. Available from: https://zqb.cyol.com/content/201001/13/content_3036671.htm.
- 8 Annas GJ. Genetic prophecy and genetic privacy--can we prevent the dream from becoming a nightmare?. American Journal of Public Health. 1995 Sep;85(9):1196-7.; Weimer DL, Vining AR. Policy analysis: Concepts and practice. Routledge; 2017 Mar 31.; Gostin LO. Genetic privacy. Journal of Law, Medicine & Ethics. 1995 Dec;23(4):320-30.
- 9 Wolf SM. Beyond "genetic discrimination": toward the broader harm of geneticism. Journal of Law, Medicine & Ethics. 1995 Dec;23(4):345-53.; Wolf SM, Kahn JP. Genetic testing and the future of disability insurance: ethics, law & policy. Journal of Law, Medicine & Ethics. 2007 Jul;35(S2):6-32.
- 10 Chapman CR, Mehta KS, Parent B, Caplan AL. Genetic discrimination: emerging ethical challenges in the context of advancing technology. Journal of Law and the Biosciences. 2020 Jan;7(1):lsz016.
- 11 Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In Artificial intelligence in healthcare 2020 Jan 1 (pp. 295-336). Academic Press.
- 12 Chapman CR. Ethical, legal, and social implications of genetic risk prediction for multifactorial disease: a narrative review identifying concerns about interpretation and use of polygenic scores. Journal of Community Genetics. 2023 Oct;14(5):441-52.
- 13 Jwa AS, Poldrack RA. Addressing privacy risk in neuroscience data: from data protection to harm prevention. Journal of Law and the Biosciences. 2022 Jul;9(2):lsac025.
- 14 Prince AE, Schwarcz D. Proxy discrimination in the age of artificial intelligence and big data. Iowa L. Rev.. 2019;105:1257.
- 15 Barocas S, Selbst AD. Big data's disparate impact. Calif. L. Rev.. 2016;104:671.; Hellman D. Measuring algorithmic fairness. Virginia Law Review. 2020 Jun 1;106(4):811-66.; Ish P, Malhotra N, Gupta N. GINA 2020: what's new and why?. Journal of Asthma. 2021 Oct 3;58(10):1273-7.; Adams-Prassl J, Binns R, KellyLyth A. Directly discriminatory algorithms. The modern law review. 2023 Jan;86(1):144-75.; Schwarcz D. Healthbased proxy discrimination, artificial intelligence, and big data. Hous. J. Health L. & Pol'y. 2021;21:95.
- 16 Babic B, Gerke S, Evgeniou T, Cohen IG. Beware explanations from AI in health care. Science. 2021 Jul 16;373(6552):284-6.
- 17 Chen X. Algorithmic proxy discrimination and its regulations. Computer Law & Security Review. 2024 Sep 1;54:106021.
- 18 MacKay DJ. Information theory, inference and learning algorithms. Cambridge University Press; 2003 Sep 25.; Angelopoulos AN, Bates S, Fannjiang C, Jordan MI, Zrnic T. Prediction-powered inference. Science. 2023 Nov 10;382(6671):669-74.
- 19 See Genetic Information Nondiscrimination Act of 2008, 42 U.S.C. § 2000ff et seq. (2008)
- 20 See GINA of 2008, 42 U.S.C. § 2000ff (4) (2008).
- 21 Cofone IN. Algorithmic discrimination is an information problem. Hastings LJ. 2018;70:1389.
- 22 Packin NG, Lev-Aretz Y. Learning algorithms and discrimination. In Research handbook on the law of artificial intelligence 2018 Dec 28 (pp. 88-113). Edward Elgar Publishing.
- 23 Weerts H, Kelly-Lyth A, Binns R, Adams-Prassl J. Unlawful proxy discrimination: a framework for challenging inherently discriminatory algorithms. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency 2024 Jun 3 (pp. 1850-1860).
- 24 Conley D, Bennett NG. Is biology destiny? Birth weight and life chances. American Sociological Review. 2000 Jun;65(3):458-67.
- 25 Schwalbe N, Wahl B. Artificial intelligence and the future of global health. The Lancet. 2020 May 16;395(10236):1579-86.
- 26 Rigotti C, Fosch-Villaronga E. Fairness, AI & recruitment. Computer Law & Security Review. 2024 Jul 1;53:105966.
- 27 Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as a medical device. NPJ digital medicine. 2020 Apr 7;3(1):53.
- 28 Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In Artificial intelligence in healthcare 2020 Jan 1:295-336. Academic Press.
- 29 Von Eschenbach WJ. Transparency and the black box problem: Why we do not trust AI. Philosophy & Technology. 2021 Dec;34(4):1607-22.
- 30 McQuillan D. Resisting AI: An anti-fascist approach to artificial intelligence. Policy Press; 2022 Jul 15.
- 31 Ehrlich I. The deterrent effect of criminal law enforcement. The Journal of Legal Studies. 1972 Jun 1;1(2):259-76.

- 32 Genetic Information Nondiscrimination Act of 2008, 42 U.S.C. §§ 2000ff-1, 300gg-53 (2008).
- 33 Prince AE, Roche MI. Genetic information, non-discrimination, and privacy protections in genetic counseling practice. *Journal of genetic counseling*. 2014 Dec;23(6):891-902.
- 34 Erel I, Myers SC, Read Jr JA. A theory of risk capital. *Journal of Financial Economics*. 2015 Dec 1;118(3):620-35.
- 35 Heinrichs B. Discrimination in the age of artificial intelligence. *AI & society*. 2022 Mar;37(1):143-54.
- 36 Benenson FC. Probability, objectivity and evidence. Taylor & Francis; 2025 Aug 29.; Joh EE. Feeding the machine: Policing, crime data, & algorithms. *Wm. & Mary Bill Rts. J.*. 2017;26:287.
- 37 Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* 2022 May 12 (pp. 296-299). Auerbach Publications.
- 38 Tencent Research Institute, CAICT, Tencent AI Lab, Tencent open platform. Invisible Injustice. In *Artificial Intelligence: A National Strategic Initiative* 2020 Oct 2 (pp. 199-208). Singapore: Springer Singapore.; Raposo VL. The fifty shades of black: about black box AI and explainability in healthcare. *Medical Law Review*. 2025;33(1):fwaf005.
- 39 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019 Oct 25;366(6464):447-53.
- 40 S. Health AI for good rather than evil? The need for a new regulatory framework for AI-based medical devices. *Yale J. Health Pol'y L. & Ethics*. 2021;20:432.
- 41 Price WN, Kaminski ME, Minssen T, Spector-Bagdady K. Shadow health records meet new data privacy laws. *Science*. 2019 Feb 1;363(6426):448-50.
- 42 Nachbar TB. Algorithmic fairness, algorithmic discrimination. *Fla. St. UL Rev.*. 2020;48:509.
- 43 Castro C, Loi M. The fair chances in algorithmic fairness: a response to Holm. *Res Publica*. 2023;29:331.
- 44 Van Laar C, Levin S. The experience of stigma: Individual, interpersonal, and situational influences. In *Stigma and Group Inequality* 2006 Aug 15 (pp. 15-32). Psychology Press.
- 45 Link BG, Phelan JC, Hatzenbuehler ML. Stigma and social inequality. In *Handbook of the social psychology of inequality* 2014 Aug 19 (pp. 49-64). Dordrecht: Springer Netherlands.
- 46 Cohen IG. Informed consent and medical artificial intelligence: what to tell the patient?. *Geo. LJ*. 2019;108:1425.
- 47 Geiger RS. Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture. *Big Data & Society*. 2017 Sep;4(2):2053951717730735.
- 48 Cohen IG, Slotte A. Artificial intelligence and the law of informed consent. *Research handbook on health, AI and the law*. 2024 Jul 16:167-82.
- 49 Apple, *User Privacy and Data Use*, <https://developer.apple.com/app-store/user-privacy-and-data-use/> (last visited Apr. 20, 2026). The author further confirmed by telephone with Apple's official customer service that the relevant content may be cited and used for academic research purposes.
- 50 Easterbrook FH. Predatory strategies and counterstrategies. *U. Chi. L. Rev.*. 1981;48:263.
- 51 Liberto H. Exploitation and the vulnerability clause. *Ethical Theory and Moral Practice*. 2014 Aug;17(4):619-29.
- 52 See Personal Information Protection Law of PRC, Article 28.
- 53 Kuttner R. Market-based failure—a second opinion on US health care costs. *New England Journal of Medicine*. 2008 Feb 7;358(6):549-51.
- 54 Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics*. 2018 Feb;26(2):149-56.
- 55 See Personal Information Protection Law of PRC, *supra* note 51, Article 28.
- 56 See Personal Information Protection Law of PRC, Article 73.
- 57 Eyer K. The but-for theory of anti-discrimination law. *Virginia Law Review*. 2021 Dec 1;107(8):1621710.
- 58 Sessa F, Chisari M, Esposito M, Karaboue MA, Salerno M, Cocimano G. Ethical, legal and social implications (ELSI) regarding forensic genetic investigations (FGIs). *Journal of Academic Ethics*. 2025 Sep;23(3):617-37.
- 59 Koshiyama A, Kazim E, Treleaven P, Rai P, Szpruch L, Pavey G, Ahamat G, Leutner F, Goebel R, Knight A, Adams J. Towards algorithm auditing. See <https://papers.ssrn.com/sol3/papers.cfm>. 2025.
- 60 Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health affairs*. 2014 Jul 1;33(7):1123-31.
- 61 Maciejewski M. To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*. 2017 Mar;83(1_suppl):120-35.

- 62 O'Brien JT, Nelson C. Assessing the risks posed by the convergence of artificial intelligence and biotechnology. *Health security*. 2020 Jun 1;18(3):219-27.
- 63 Mergel I. Social affordances of agile governance. *Public Administration Review*. 2024 Sep;84(5):932-47.
- 64 Wang R, Read I. An Agile Approach: Understanding China's AI Governance Framework. *Lexology* [Internet]. 2025 Nov 5 [cited 2026 Mar 22]. Available from: <https://www.lexology.com/library/detail.aspx?g=ef8efea8-05b1-4bf2-8fd5-08ac531bfd7d>
- 65 Schjoedt L, Sangboon K, Bird B. Control variables: Use, misuse, and recommended use. In *Handbook of Research Methods and Applications in Entrepreneurship and Small Business*, Second Edition 2025 Jan 7 (pp. 186-204). Edward Elgar Publishing.
- 66 Mohapatra P. Intelligent Test Prioritization in Risk-Based Testing Using Artificial Intelligence. *Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle*. 2025 Jul 27;4:74.
- 67 Corfmat M, Martineau JT, Régis C. High-reward, high-risk technologies? An ethical and legal account of AI development in healthcare. *BMC Medical Ethics*. 2025 Jan 15;26(1):4.
- 68 Al-Omush A, Almasarwah A, Al-Wreikat A. Artificial intelligence in financial auditing: redefining accuracy and transparency in assurance services. *EDPACS*. 2025 Jun 3;70(6):1-20.
- 69 Cheong BC. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*. 2024 Jul 3;6:1421273.
- 70 Sangiuliano A. Against moralism in anti-discrimination law. *University of Toronto Law Journal*. 2023 Oct 1;73(4):467-98.
- 71 Wachter S. The theory of artificial immutability: Protecting algorithmic groups under antidiscrimination law. *Tulane Law Review* 2022;97:149.
- 72 Thornton M. Anti-discrimination legislation in Australia: A reckoning. *International Journal of Discrimination and the Law*. 2025 Dec;25(4):388-413.
- 73 Supreme People's Court of the People's Republic of China. Civil Code of the People's Republic of China: Article 1167 [Internet]. Faxin. [cited 2026 May 22]. Available from: <https://xcx2019.faxin.cn/m/v5/web/twsy/TwsyContent.aspx?isapp=1&gid=A290592&tiao=1167>