information

MDPI

Article

Systemic Data Bias in Real-World AI Systems: Technical Failures, Legal Gaps, and the Limits of the EU AI Act

Theodoros Falelakis [1], Asimina Dimara [2,*] and Christos-Nikolaos Anagnostopoulos [2,*]

[1] Faculty of Law, Aristotle Universtiy of Thessaloniki, 541 24 Thessaloniki, Greece; theofale@law.auth.gr
[2] Department of Cultural Technology and Communication, University of the Aegean, 811 00 Mitilini, Greece
* Correspondence: dimara@aegean.gr (A.D.); canag@aegean.gr (C.-N.A.)

Abstract

Systemic data bias constitutes a major source of failure in real-world AI systems and represents a regulatory challenge that remains insufficiently addressed by existing legal frameworks, including the EU Artificial Intelligence Act. Although the AI Act introduces a comprehensive risk-based regulatory regime, it does not adequately capture how bias originates, propagates, and manifests across the AI lifecycle. This paper examines systemic data bias through a legal-technical lifecycle analysis that maps recurring bias mechanisms, from data collection and annotation to model training, evaluation, and deployment, to the regulatory control points established under the EU AI Act. Drawing on cross-sectoral examples from employment screening, credit scoring, healthcare risk prediction, biometric identification, and autonomous systems, the analysis demonstrates how technical bias mechanisms translate into systemic governance and accountability challenges. The findings reveal persistent regulatory gaps, including limited auditability of training datasets, the absence of mandatory fairness metrics, insufficient transparency regarding model behavior, and weak mechanisms for post-deployment monitoring and accountability. These results highlight a structural misalignment between lifecycle-based bias dynamics and the Act's category-driven compliance framework. The paper argues that addressing systemic bias requires a governance approach that integrates technical bias mitigation with legal oversight across the full AI lifecycle rather than relying primarily on post hoc regulatory controls.

Keywords: systemic data bias; AI governance; EU AI Act; fairness metrics; algorithmic accountability; socio-technical systems

Check for updates

Academic Editor: Sheng Gao

Received: 30 January 2026
Revised: 4 March 2026
Accepted: 25 March 2026
Published: 27 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Artificial intelligence (AI) emerged as a scientific field in the 1950s, initially focused on symbolic reasoning, logic systems, and the formalization of human cognition [1]. The field was formally established in 1956 at the Dartmouth Conference, which is widely regarded as the birth of AI as a distinct area of research [1]. One of the first instances in which AI visibly interacted with humans in a real-world setting occurred through game systems, most notably the chess match between IBM's Deep Blue and Garry Kasparov in 1997, which demonstrated machine superiority in narrowly defined, rule-based environments [2]. From the early 1980s onward, AI systems gradually moved beyond games into practical decision-making applications, including expert systems and automated agents used in domains such as credit scoring, employment screening, healthcare risk assessment, and operational optimization [3]. These systems increasingly influenced high-stakes decisions

affecting individuals and organizations and became embedded in everyday socio-economic processes for several decades.

However, while AI systems had been widely deployed in real-world decision-making long before, comprehensive legal regulation emerged only recently. In the European Union (EU), the Artificial Intelligence Act (AI Act), proposed in 2021 and adopted in 2024, constitutes the first horizontal legal framework specifically targeting AI systems [4]. This development highlights a substantial temporal gap of nearly four decades between the technological maturation and societal deployment of AI and the establishment of binding regulatory oversight. In contrast, the United States has historically relied on pre-existing sector-specific laws originally designed for human and rule-based decision-making, such as the Fair Credit Reporting Act (1970), the Equal Credit Opportunity Act (1974), and the Privacy Act (1974). These legal instruments were later applied to automated and AI-driven systems through judicial interpretation and regulatory enforcement rather than through AI-specific legislation [5].

At the same time, it is important to clarify that the European Union was not operating in a regulatory vacuum prior to the AI Act. Long before the adoption of a dedicated horizontal AI framework, several sector-specific and fundamental rights instruments, originally drafted for human or conventional automated decision-making, were interpreted and enforced so as to cover also algorithmic and AI-based systems, precisely in the absence of a unified and expressly AI-specific regulatory framework. Most prominently, the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) established rules on automated individual decision-making, including profiling (Article 22), transparency obligations (Articles 13–15), data protection by design and by default (Article 25), and fairness and lawfulness principles (Article 5). In parallel, EU anti-discrimination law, including Directive 2000/43/EC (Racial Equality Directive), Directive 2000/78/EC (Employment Equality Directive), and Directive (EU) 2024/2831 on improving working conditions in platform work, has provided a substantive framework for challenging discriminatory algorithmic outcomes in areas such as employment and access to services. In financial services, supervisory and regulatory instruments have also contributed to the scrutiny of automated credit-scoring systems. In particular, the Article 29 Data Protection Working Party Guidelines on Automated Individual Decision-Making and Profiling under Regulation (EU) 2016/679, Directive (EU) 2023/2225 on credit agreements for consumers (recasting the Consumer Credit Directive), and more recently Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (DORA) have introduced obligations relating to automated decision-making, governance, data protection, and ICT risk management, which are applicable to algorithmic creditworthiness assessments. Although not AI-specific instruments, these frameworks have been used in practice to oversee and constrain the deployment of credit-scoring algorithms. Furthermore, product safety and liability regimes, including the Product Liability Directive (85/374/EEC), have offered additional provisions for redress where AI-enabled products cause harm.

Accordingly, much like the United States, the EU relied for years on the adaptive interpretation of pre-existing sectoral legislation to govern AI-related risks; the AI Act therefore does not replace an absence of regulation, but rather it forms a landscape into a coherent, risk-based horizontal regime. However, it should also be noted that the AI Act does not establish uniform obligations for all AI systems. Its most stringent compliance requirements apply only to high-risk AI systems as exhaustively listed in Annex III and related provisions. As a result, only a relatively limited subset of AI applications, compared to the vast spectrum of everyday and foreseeable future AI uses, falls within the core regulatory intensity of the Act, while the majority of low-risk and minimal-risk systems remain subject mainly to general EU law, national law, and, in certain contexts, soft-law

instruments such as guidelines and voluntary codes of conduct. This structural limitation and its practical implications for the overall regulatory coverage of AI systems will be examined in greater detail in the following sections of this research.

Furthermore, in real-world AI decision-making, numerous high-impact failures have revealed severe forms of bias that were neither anticipated nor effectively addressed by existing legal frameworks. Well-documented cases include automated hiring systems that systematically disadvantaged female candidates due to biased training data [6], credit scoring models that assigned lower credit limits to women and minority groups despite equivalent financial profiles [7], and healthcare risk prediction systems that underestimated the medical needs of marginalized populations by relying on cost-based proxies rather than clinical indicators [8]. Similar failures have been observed in biometric identification and autonomous systems, where biased datasets and evaluation practices led to disproportionate error rates across demographic groups, creating tangible risks to safety and fundamental rights [9]. These cases demonstrate that bias in AI decision-making is not an isolated technical anomaly but a recurring and systemic phenomenon with material legal and societal consequences [10]. Crucially, many of these harms emerged in systems that formally complied with existing regulatory requirements, exposing a gap between legal oversight mechanisms and the technical realities of how AI systems are designed, trained, and deployed [11].

Within this context, this paper examines systemic data bias as a primary source of failure in real-world AI systems and as a regulatory challenge that remains insufficiently addressed by current legal frameworks, including the EU AI Act. Rather than treating bias as a narrow technical defect, the paper conceptualizes it as a structural feature of socio-technical systems that originates and propagates across the AI lifecycle, from data collection and annotation to model training, evaluation, and deployment. This framing reflects the broader insight that bias often operates beneath formally articulated reasons, creating *a disparity between the underlying factors that truly influence the decision-making process and the apparent adherence to legal reasons* [1]. The novelty of this study lies in its integrated treatment of technical bias mechanisms and legal governance, bridging domains that are typically analyzed in isolation. Through a sectoral, case-based legal and technical analysis spanning employment, credit scoring, healthcare risk prediction, biometric identification, and autonomous systems, the study demonstrates how recurring technical bias mechanisms translate into large-scale operational failures and legal accountability gaps. Unlike existing work that focuses either on technical mitigation or regulatory design in abstraction, this paper maps real-world bias failures to the risk-based structure of the EU AI Act. By mapping these failures against the EU AI Act's regulatory architecture, the paper identifies persistent blind spots in auditability, fairness evaluation, continuous monitoring, and liability attribution. In line with prior findings that data-driven systems can be used *to identify instances of bias, arbitrariness, or disproportionality,* the analysis shows that such identification remains largely optional rather than structurally embedded within binding governance obligations [12]. Based on these findings, it proposes a unified technical and legal governance approach that integrates bias mitigation into compliance by design and reframes legal frameworks as co-designed components of AI systems rather than post hoc regulatory constraints.

Unlike much of the existing literature, which typically examines either technical bias mitigation techniques or regulatory design in abstraction, this study introduces a lifecycle-to-regulatory mapping framework that systematically links technical bias mechanisms across the AI lifecycle to the concrete control points established under the EU AI Act. By synthesizing cross-sectoral examples from multiple high-impact domains, the analysis also develops a structured identification of recurring governance gaps, including dataset

auditability limitations, the absence of mandatory fairness metrics, weaknesses in post-deployment monitoring, and ambiguities in liability attribution. In order to structure the analysis, the study is guided by three interrelated research questions:

1. how and at which stages of the AI lifecycle systemic data bias emerges and propagates;
2. how these bias mechanisms interact with the regulatory control points established under the EU AI Act; and
3. which governance gaps or blind spots persist when lifecycle bias dynamics are examined in relation to the Act's risk-based regulatory framework.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature on failures of AI systems from technical, socio-technical, and legal perspectives, alongside existing scholarship on the EU AI Act and its risk-based regulatory approach. Section 3 establishes the conceptual foundations of the study, framing data bias as a systemic socio-technical phenomenon and mapping bias mechanisms to regulatory control points across the AI lifecycle. Section 4 outlines the methodological approach and presents the sectoral, case-based legal and technical analysis applied across multiple high-impact domains. Section 5 synthesizes the cross-sectoral findings, identifying recurring technical patterns and structural governance gaps that enable the persistence of biased AI systems despite formal regulatory compliance. Finally, the last section concludes the paper by summarizing the main findings and clarifying their implications for the limits of risk-based AI governance.

2. Literature Review

This section reviews the relevant literature on failures of AI systems and on the regulatory approaches developed to address them, with a particular focus on the EU AI Act. The review highlights limitations and blind spots in existing approaches that motivate the integrated legal–technical analysis pursued in this paper.

2.1. Failures of AI Systems: Technical, Socio-Technical, and Legal Perspectives

This section reviews the existing literature on AI system failures from technical, socio-technical, and legal perspectives. It examines how each body of scholarship conceptualizes the sources and consequences of bias in AI systems and highlights the differing analytical lenses through which failures are understood. By structuring the review across these perspectives, the section sets the foundation for identifying gaps and fragmentation in current approaches.

2.1.1. Technical Failure

The literature on AI failures has documented how bias arises from data-related and model-specific mechanisms, including unrepresentative or imbalanced datasets, biased labeling practices [13], the use of proxy variables correlated with protected attributes [14], and the selection of performance metrics that optimize aggregate accuracy at the expense of subgroup fairness [15]. While numerous statistical fairness metrics have been proposed in the technical literature, the EU AI Act does not mandate the use of any specific fairness evaluation framework, leaving significant discretion to providers in determining how bias is assessed. Moreover, many statistical fairness metrics proposed in the technical literature have been developed primarily within U.S. regulatory and legal contexts, and their direct application within EU governance frameworks may raise compatibility questions with the doctrinal structure of EU equality and anti-discrimination law.

A plethora of work focuses on identifying statistical disparities across demographic groups and proposing mitigation techniques such as data rebalancing, fairness-aware learning objectives, post-processing adjustments, and explainability tools [16]. Within this

strand, bias is mainly framed as an engineering problem that can be detected, quantified, and mitigated through improved data practices and model re-design [17], while these contributions provide critical insights into the technical origins of biased behavior, they typically abstract away from the broader socio-institutional context in which AI systems operate and do not engage with questions of legal accountability, regulatory compliance, or governance obligations [18]. As a result, technical failure is often treated as a localized defect within the model or dataset rather than as a systemic risk that propagates across the AI lifecycle and intersects with binding legal frameworks [19].

2.1.2. Socio-Technical Failure

The socio-technical work on AI failures examines how biased or erroneous model outputs are amplified through human interaction, organizational practices, and institutional contexts [20]. This body of work emphasizes mechanisms such as automation bias, feedback loops, and overreliance on algorithmic recommendations, demonstrating how AI systems can reinforce existing inequalities once deployed in real-world settings [21]. Studies highlight that failures often emerge not solely from flawed models but from the interaction between technical systems and human decision-makers who defer to algorithmic outputs, sometimes despite contradictory evidence or contextual knowledge [22]. Socio-technical analyses further show how biased outcomes can become self reinforcing over time, particularly in domains such as policing, hiring, and credit allocation, where system outputs influence future data collection [23], while this literature provides valuable insights into the systemic amplification of bias after deployment, it generally treats the internal technical mechanisms of AI systems and the corresponding regulatory obligations as given, without explicitly linking socio-technical failure dynamics to lifecycle-based bias propagation or to concrete legal governance frameworks [24].

2.1.3. Legal and Regulatory Perspectives on AI Failures

Recent legal and regulatory works have increasingly examined AI failures through the lens of risk governance, accountability, and fundamental rights, particularly following the proposal of the EU AI Act in 2021 [25]. Post-2021 literature has focused on the limitations of risk-based regulatory approaches, the challenges of attributing responsibility in complex AI supply chains, and the adequacy of existing legal instruments in addressing discriminatory and biased outcomes generated by automated systems [26]. Several studies highlight that current regulatory frameworks prioritize ex ante conformity assessments, documentation requirements, and procedural compliance while offering limited mechanisms for detecting and addressing bias that emerges dynamically during deployment and use [27]. Legal analyses published after 2021 also emphasize persistent gaps in dataset auditability, enforceable fairness obligations, and continuous post-deployment monitoring, noting that many AI-related harms fall outside clear liability regimes [28]. Although this body of work provides critical insights into governance design and regulatory intent, it often treats technical bias mechanisms at a high level of abstraction and does not systematically engage with how bias originates and propagates across the AI lifecycle. As a result, legal scholarship frequently addresses AI failures at the level of normative principles and regulatory structure, without a detailed linkage to the underlying technical processes that generate biased outcomes in real-world systems [29].

As discussed in the above sections, the literature offers valuable but fragmented accounts of AI system failures. Technical studies primarily focus on bias arising from data and model design, socio-technical analyses emphasize the amplification of biased outputs through human interaction and institutional practices, and legal scholarship addresses governance, accountability, and compliance at a high level of abstraction. However, these

perspectives are typically developed in isolation, without a unified framework that explains how bias originates, propagates, and manifests across the full AI lifecycle. As a result, the mechanisms through which data bias introduced during early stages, such as data collection or annotation, translate into downstream operational failures remain insufficiently connected to regulatory oversight and legal responsibility. This fragmentation limits the capacity of existing approaches to anticipate and mitigate real-world harms, particularly when bias emerges or intensifies after deployment. Consequently, the absence of a lifecycle-based bias propagation model obscures the link between technical design choices and their legal and societal implications. Figure 1 illustrates how technical failure sources, AI bias manifestations, and systemic legal consequences are extensively discussed in the literature while remaining insufficiently connected through a lifecycle-based governance framework. In particular, early-stage data and design choices are weakly linked to downstream regulatory obligations, legal accountability, and post-deployment corrective mechanisms.

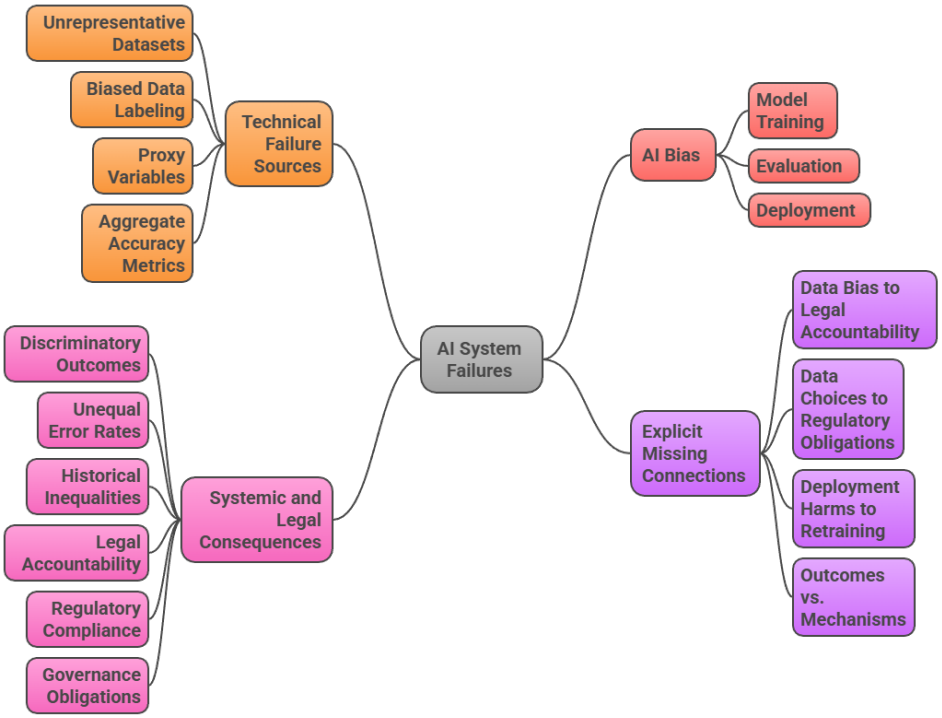


Figure 1. Fragmentation between technical sources of data bias, AI lifecycle stages, and legal accountability in AI systems, illustrating how bias originating in data and model design propagates through deployment while legal accountability mechanisms remain disconnected from these technical processes.

2.2. The EU AI Act in the Literature: Risk-Based Regulation and Its Limits

The EU AI Act's regulatory architecture contains significant gaps that compromise its foundational objectives of safeguarding fundamental rights while enabling technological innovation. Leading legal scholars and researchers have identified systemic weaknesses, regulatory blind spots, and exploitable loopholes. Although the EU AI Act claims to embody risk-based regulation, the literature demonstrates that several core elements fundamentally undermine this commitment. First, protecting fundamental rights through a risk-based approach is conceptually flawed [30]. Despite the fact that the EU AI Act speaks about risks to rights as its prime objective, with 80 mentions in the initial proposal [31], the EU AI Act expressly provides for individual rights in the two only Articles about remedies, namely the right to lodge a complaint (Art. 85) and to explanation (Art. 86). These provisions are the only ones expressly providing for individual rights and, at the same time, the only ones specifically addressing individual remedies, which constitutes a

further shortcoming of the AI Act. Another issue that is identified is that the AI Act does not provide a clear risk–benefit analysis since it answers only to the risks, missing how those risks can be considered acceptable in light of the benefits [30]. This is illustrated especially concerning health care, when the benefits of AI's use outweigh its risks [32].

The EU AI Act has also received criticism for its approach to fundamental rights protection through transparency provisions and the absence of a dedicated liability regime within the Act, thereby leaving affected individuals to rely on pre-existing traditional and insufficient liability mechanisms, which together create insurmountable practical barriers to effective redress [33]. The transparency framework comprises a complex architecture of exemptions, exceptions, derogations, and restrictions that lack a coherent pro-disclosure foundation, rendering the scope of transparency obligations fragmented and context dependent without systematic guidance [34]. Critically, individuals seeking redress face dual evidentiary obstacles: they must first achieve awareness of AI-induced harms, particularly challenging for immaterial injuries such as algorithmic discrimination in advertising, credit, or employment contexts, and subsequently satisfy nearly insurmountable standards of proof, especially when lacking technical expertise or legal representation [35]. This is mostly because, in the absence of a specific liability framework under the AI Act, victims must rely on traditional fault-based regimes ill-suited to the opacity and technical complexity of AI systems, while the regulatory processes largely substituted empirical validation with the collection of views and attitudes from selected stakeholder groups, making the framework a result of political compromise rather than one grounded in robust scientific evidence or systematic statistical measurement of AI functionalities [36].

Additional regulatory deficiencies include the vague terminology governing prohibited manipulative techniques (particularly the undefined threshold of influence *beyond consciousness*), which complicates enforcement of Article 5 prohibitions [37]. For generative AI systems, the absence of mandatory continuous logging throughout the model lifecycle and the restriction of incident reporting requirements to only systemically risky models eliminates visibility into performance degradation and emergent harms among standard GPAI providers [38]. The literature converges on the view that the AI Act's risk-based architecture leaves meaningful *blind spots* where the most intrusive uses can fall outside (or be softened within) the regime. Even within the prohibited practices tier, layered carve-outs can preserve deployment in practice (e.g., the *exceptions* structure around Art. 5(1)(d) in conjunction with (2) to (7)) [39]. A widely discussed example is the effective creation of a parallel compliance framework exempting law enforcement, migration, and national security authorities from the Act's most stringent safeguards—a consequence of sustained pressure from Member States and security industry stakeholders [40]. Furthermore, the legislation explicitly excludes military applications of AI systems, despite recognizing their high-risk profile and the necessity for human control and legal accountability in such contexts [41]. Similarly, AI systems deployed for research purposes fall outside the regulatory scope [42], creating implementation uncertainties where research-derived methodologies transition to practical applications in regulated sectors such as pharmaceutical assessment. These exemptions and exclusions establish considerable blind spots where the most dangerous applications of AI technology operate with minimal oversight, directly undermining the Act's protective mission. These convergent limitations establish a regulatory system where individual rights enforcement depends on conditions rarely satisfied, and technological oversight remains inherently reactive rather than preventive.

The literature has also identified that the regulatory framework suffers from profound fragmentation across the EU's heterogeneous legal landscape, with Member States operating within distinct legal traditions and administrative capacities [43–45]. This fragmentation manifests across three critical dimensions: disparate enforcement capabilities

among national authorities, divergent interpretations of regulatory requirements, and inconsistent implementation timelines, conditions that create substantial compliance barriers for organizations operating across multiple EU jurisdictions. The conformity assessment mechanism, designed as the Act's primary compliance verification instrument, exhibits significant weaknesses [35]. For high-risk AI systems listed in Annex III (covering education, workplace, and financial services), assessments are self-certified by providers without public disclosure, whereas third-party assessments apply only to AI systems functioning as safety components in regulated products. This bifurcated approach compounds with the governance framework's insufficient specification regarding mechanisms to ensure impartiality and prevent conflicts of interest among notifying authorities, lacking explicit assignment of enforcement responsibility [46]. The net result is a system where market participants possess substantial discretion in demonstrating compliance, accountability mechanisms remain opaque, and oversight mechanisms depend on inconsistently resourced national regulators.

3. Conceptual Foundations

This section establishes the conceptual framework that informs the analysis of systemic data bias in real-world AI systems. Drawing on insights from the literature on AI failures and the regulatory limitations of the EU Artificial Intelligence Act, it frames data bias as a socio-technical and lifecycle-wide phenomenon rather than a localized technical defect. The section clarifies key conceptual distinctions and subsequently operationalizes this framework by mapping recurring technical bias mechanisms across the AI lifecycle to the regulatory control points established under the EU AI Act, thereby providing the analytical basis for the cross-sectoral and governance analysis that follows.

3.1. Data Bias as a Socio-Technical Phenomenon

Data bias refers to distortions introduced during data collection, selection, labeling, and preprocessing. Algorithmic bias arises from model design choices, including learning objectives, feature representations, and evaluation metrics. AI bias refers to the observable outcomes of deployed systems, emerging from the interaction of data bias, algorithmic bias, and the socio-institutional environment in which the system operates. In this sense, systemic bias reflects not only technical distortions but also feedback loops and institutional practices that reinforce biased patterns over time. Data bias is frequently discussed as a technical flaw within datasets or AI models; however, this framing obscures its broader systemic nature. To avoid conceptual ambiguity, it is important to distinguish between related but distinct forms of bias in AI systems. Data bias refers to systematic distortions introduced during data collection, selection, labeling, and preprocessing, often reflecting historical inequalities, measurement limitations, or sampling imbalances [47]. Algorithmic bias arises from model design choices, including learning objectives, feature representations, optimization strategies, and evaluation metrics that may unevenly distribute errors across subgroups [48]. In contrast, AI bias describes the observable, real-world behavior of deployed systems, emerging from the interaction of data bias, algorithmic bias, and the socio-institutional context in which AI systems operate [49]. This distinction clarifies that biased outcomes are rarely attributable to a single technical component but rather to interdependent processes spanning the AI system as a whole.

Viewed through a lifecycle perspective, data bias originates early and propagates through successive stages of AI development and deployment [47]. Bias introduced during data collection or annotation is often amplified during model training, reinforced through evaluation practices that prioritize aggregate performance, and ultimately manifested during deployment in the form of unequal error rates, discriminatory outcomes, or reduced

reliability for specific populations [47]. Once deployed, AI systems can further influence the data they generate, creating feedback loops that entrench and normalize biased patterns over time. This lifecycle-based understanding highlights that bias is not a static defect that can be isolated and removed at a single point but a dynamic process that evolves across technical and organizational stages.

Consequently, purely technical mitigation strategies, while necessary, are insufficient on their own to address systemic data bias [48]. Techniques such as rebalancing datasets, modifying loss functions, or adjusting decision thresholds can reduce measurable disparities, but they do not resolve questions of accountability, oversight, or responsibility when biased outcomes persist or re-emerge after deployment [50]. Therefore, without governance mechanisms that align technical design choices with regulatory obligations, monitoring requirements, and legal accountability structures, bias mitigation remains fragmented and fragile. Framing data bias as a socio-technical phenomenon therefore provides the conceptual foundation for integrating technical interventions with governance and regulatory design, a linkage that is essential for addressing bias in real-world AI systems [51].

As depicted in Figure 2 data bias originates during early lifecycle stages, is shaped by algorithmic design choices, and manifests as biased system behavior during deployment. Socio-technical dynamics and feedback loops further reinforce bias over time. The absence of governance mechanisms spanning the full lifecycle highlights why technical mitigation alone is insufficient without integrated legal and regulatory oversight.

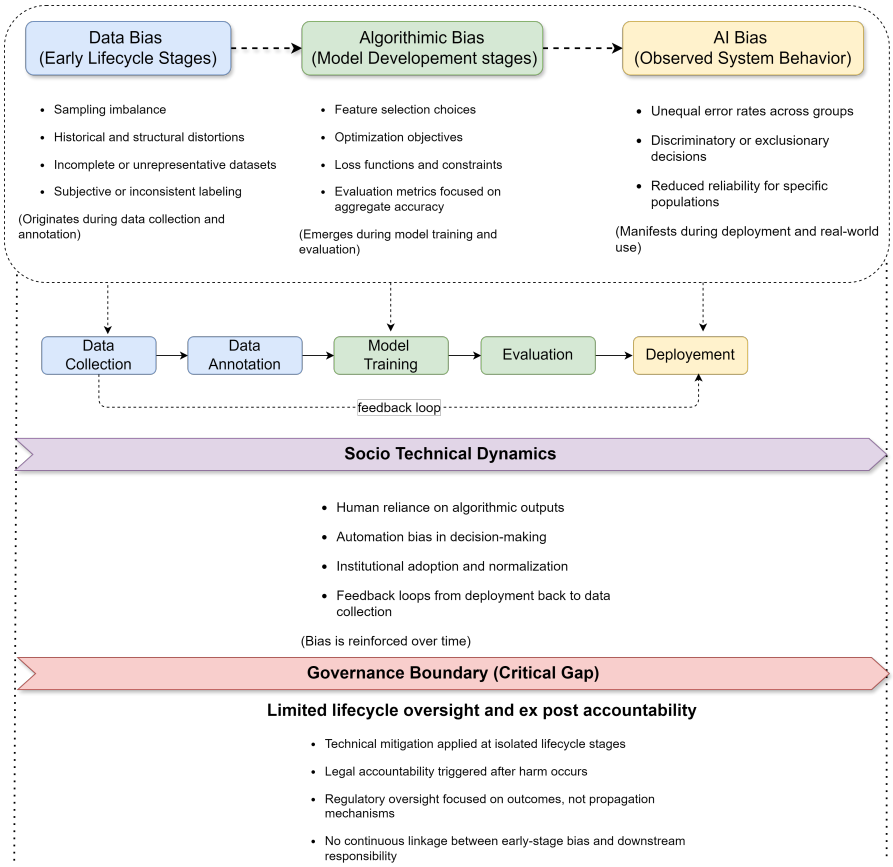


Figure 2. Data bias as a socio-technical phenomenon across the AI lifecycle, illustrating how bias emerges during early data and model development stages, propagates through deployment via socio-technical feedback loops, and remains insufficiently addressed by governance mechanisms that focus primarily on post hoc accountability.

3.2. Mapping Bias Mechanisms to Regulatory Control Points

Building on the analysis of the EU AI Act presented in Section 2.2 and the conceptual framing of data bias developed in Section Data Bias as a Socio-Technical Phenomenon (Section 3.1), this subsection examines how technical bias mechanisms align with the regulatory control points established under the AI Act, while existing scholarship has extensively analyzed the Act's risk-based architecture and governance objectives, and technical literature has documented bias mechanisms across the AI lifecycle, these strands are rarely examined in direct relation to one another. As a result, the interaction between lifecycle-based bias dynamics and regulatory oversight remains insufficiently articulated.

From a regulatory perspective, the AI Act introduces a set of ex ante and ex post obligations intended to manage risks associated with high-risk AI systems, including requirements related to data governance, risk management, human oversight, transparency, and post-market monitoring. However, as discussed in the literature, these obligations are largely formulated at a procedural and system-level abstraction, without explicit reference to the mechanisms through which bias is introduced, propagated, and reinforced across the AI pipeline. Conversely, socio-technical analyses of data bias emphasize that bias emerges early, evolves dynamically, and is shaped by feedback effects during deployment, often in ways that are not readily captured by static compliance checks or documentation-based assessments.

To bridge this gap, the present analysis adopts a lifecycle oriented mapping approach that systematically aligns recurring technical bias mechanisms with the corresponding regulatory control points of the AI Act. Rather than assessing regulatory provisions in isolation, this approach examines where specific bias mechanisms intersect with, fall outside, or are only indirectly addressed by existing legal obligations. In doing so, it becomes possible to distinguish between areas where regulatory intent and technical risk are broadly aligned and areas where structural mismatches persist.

Table 1 translates the conceptual framework into a structured alignment by linking stages of the AI lifecycle, dominant technical bias mechanisms, observed failure patterns, and the relevant provisions of the AI Act. While the lifecycle stages and technical bias dynamics described in the table apply to AI systems more generally, the regulatory control points referenced (e.g., Arts. 9–15, 61) become legally binding primarily where systems are classified as high-risk under the AI Act. The table therefore illustrates the structural alignment between bias mechanisms and the Act's control architecture, rather than suggesting that all AI systems are uniformly subject to these obligations. The table further identifies the limitations associated with each control point, highlighting where regulatory requirements remain primarily procedural, reactive, or insufficiently specified to address bias as a lifecycle wide phenomenon. Importantly, the table does not suggest an absence of regulation but rather illustrates how bias-related risks may escape effective oversight due to misalignment between technical bias dynamics and the scope of existing regulatory mechanisms.

The mapping presented in Table 1 is based on an analytical alignment between recurring bias mechanisms identified in the technical literature and the regulatory control points introduced by the EU AI Act, examining where these mechanisms intersect with, fall outside, or are only indirectly addressed by existing legal obligations.

By making these alignments and gaps explicit, this subsection provides the analytical foundation for the subsequent examination of cross sectoral governance blind spots. The mapping clarifies why compliance with the AI Act does not necessarily translate into effective bias mitigation in practice and motivates the need for governance approaches that integrate technical bias awareness into regulatory design and enforcement across the full AI lifecycle.

Table 1. Mapping technical bias mechanisms to regulatory control points under the EU AI Act.

AI Lifecycle Stage	Technical Bias Mechanism	Observed Failure Pattern	Relevant AI Act Control Points	Regulatory Gap/Limitation
Data Collection	Use of unrepresentative or historically biased datasets	Systematic exclusion or misrepresentation of specific population groups	Art. 10 (Data and data governance)	Representativeness and governance requirements are largely procedural, with no obligation for independent or outcome-oriented dataset audits
Data Annotation	Labels derived from past human decisions treated as ground truth	Reinforcement of historical discriminatory practices	Art. 10(2) 10(3) (Training, validation, and testing data relevance)	No explicit requirement to assess bias embedded in labeling practices or decision provenance
Problem Formulation	Selection of target variables misaligned with social objectives	Structural bias encoded in prediction goals	Art. 9 (Risk management system)	No obligation to justify or document the normative assumptions underlying target variable selection
Feature Construction	Use of proxy variables correlated with protected characteristics	Indirect discrimination despite formal exclusion of sensitive attributes	Recitals on indirect discrimination; Annex III	Proxy-based bias addressed indirectly through anti-discrimination principles, without explicit ex ante technical constraints
Model Training	Optimization objectives prioritizing aggregate accuracy	Unequal error rates across demographic subgroups	Art. 15 (Accuracy, robustness, and cybersecurity)	Fairness metrics and subgroup performance requirements are not mandated
Model Evaluation	Validation based on global performance metrics	Bias undetected during testing, emerges post-deployment	Art. 9 (Risk management system)	Risk management focuses on documentation rather than empirical bias detection across subgroups
Deployment	Automated decision-making at scale with limited intervention	Discriminatory outcomes normalized through operational use	Art. 14 (Human oversight)	Oversight requirements lack operationalized criteria for effective and timely human intervention
Post-Deployment	Feedback loops reinforcing biased outcomes	Bias intensifies and stabilizes over time	Art. 61 (Post-market monitoring)	Monitoring obligations are reactive and do not require continuous reassessment of bias dynamics
Accountability	Limited transparency and explainability of model behavior	Difficulty contesting biased decisions	Art. 13 (Transparency), Arts. 85–86 (Rights and remedies)	High evidentiary burden on affected individuals and weak enforceability of corrective mechanisms

Data governance and bias under Article 10. Where AI systems qualify as high risk, the data governance obligations of Article 10 are central to addressing the bias mechanisms identified above. Providers must ensure that training, validation, and testing datasets are relevant, sufficiently representative, and, to the best extent possible, free of errors and complete for the intended purpose, with documented practices covering data origin, preparation, assumptions, and statistical properties. Crucially, Article 10(2)(f)–(h) requires an examination of likely biases with negative impacts on fundamental rights or discriminatory effects, particularly where outputs influence future inputs, and mandates appropriate measures to detect, prevent, and mitigate such biases. Article 10(5) further permits, under strict safeguards, the exceptional processing of special categories of personal data solely to detect and correct bias where this cannot be effectively achieved by other means. In the credit context, these provisions directly engage the risk that proxy variables (e.g., postcode, spending patterns, and employment stability) encode structural inequalities, obliging providers to confront rather than merely document such distortions.

Regulatory blind spots Although Article 10 of the AI Act establishes an extensive data governance framework for high risk AI systems including documentation of design choices, data collection and annotation practices, the formulation of assumptions, the examination of potential bias, and the identification of shortcomings and mitigation measures, it remains normatively underdetermined as to how bias is to be defined, measured, and corrected in practice, while Articles 10(2)–(3) require datasets to be relevant, representative, and, to the best extent possible, free of errors and complete, the Regulation provides no guidance on acceptable levels of bias, legally compliant mitigation strategies, or the appropriate response where bias cannot be detected or effectively mitigated. Nor does it articulate

how to evaluate different forms of bias, such as historically embedded structural bias, bias related to *ground truth,* or culturally contingent conceptions of fairness.

As *Wachter* further argues, the choice of fairness metrics is not normatively neutral [51]. Most widely used bias tests have been developed in the United States under fundamentally different anti discrimination regimes and often fail to meet the standards of European equality law. The operationalization of *fairness* through such metrics risks producing legally non-compliant outcomes in the EU, even where providers follow state-of-the-art technical practices. Moreover, certain group fairness approaches may achieve parity only through *leveling down* [52], worsening outcomes for all groups rather than improving the position of disadvantaged populations, an outcome that is ethically troubling and difficult to reconcile with the teleology of EU non-discrimination law.

Consequently, although Article 10 formally mandates the identification and mitigation of bias, it delegates the most normatively consequential choices—what constitutes legally relevant bias, which metrics are acceptable, and how competing notions of fairness should be reconciled—to harmonized standards and private technical practice. In a sector such as credit and financial scoring, where discriminatory effects are often subtle, proxy based, and cumulative, this regulatory design creates a structural blind spot: systems may satisfy the procedural requirements of data governance while continuing to generate disparate access to credit in ways that are difficult to evaluate against EU equality norms. In *Wachter*'s terms, the regulation treats bias as a technical quality problem rather than as a fundamentally legal and normative question, thereby risking a gap between formal compliance and substantive non-discrimination.

4. Methodology: Sectoral Case-Based Legal–Technical Analysis

This section outlines the methodological approach adopted in the study, combining a sectoral, case-based analysis with a legal and technical examination of real-world AI systems. The methodology is designed to capture how data bias emerges and propagates across different stages of the AI lifecycle while enabling systematic comparison across high-impact domains. By integrating technical failure analysis with regulatory assessment, the approach supports the identification of recurring bias mechanisms and governance gaps across sectors.

4.1. Sector Selection Rationale

This study examines systemic data bias across five main application domains in which AI systems are widely deployed to support or automate high-stakes decision-making: employment and hiring, credit and financial scoring, healthcare risk prediction, biometric identification, and autonomous systems. The sectoral sections that follow are not presented as formal case studies but as analytical illustrations used to examine how recurring bias mechanisms and governance gaps manifest across different AI application domains. These sectors were selected because they combine intensive data processing, consequential decision outcomes, and sustained human–AI interaction while also occupying a central position in contemporary regulatory and governance debates. Together, they provide a representative cross-section of real-world AI deployment contexts in which bias related failures have been repeatedly documented and publicly contested.

From a technical standpoint, the selected sectors encompass a diverse set of AI system architectures, data modalities, and operational contexts. They include systems based on structured tabular data (e.g., credit scoring and hiring), predictive risk models relying on proxy variables (e.g., healthcare risk prediction), perception-driven systems using image and sensor data (e.g., biometric identification and autonomous systems), and hybrid decision support systems that combine automated inference with human judg-

ment. This diversity enables the analysis to identify recurring bias mechanisms that are not confined to a single model type or application domain but instead emerge from shared patterns in data collection, labeling, feature construction, evaluation practices, and deployment environments.

From a socio-technical perspective, these sectors are characterized by strong institutional embedding and high levels of reliance on algorithmic outputs. Decisions supported by AI systems in employment, finance, and healthcare shape organizational practices and future data generation, creating feedback loops that can reinforce initial biases over time. In biometric identification and autonomous systems, biased performance can translate directly into unequal treatment, exclusion, or safety risks. The selected domains, therefore, provide an appropriate context for examining how technical bias mechanisms are amplified through human reliance, organizational routines, and broader social structures.

From a regulatory perspective, the chosen sectors closely align with the EU Artificial Intelligence Act's risk-based framework, encompassing applications that are explicitly classified as high risk as well as systems operating near regulatory boundaries. Employment, creditworthiness assessment, healthcare, biometric identification, and safety-critical autonomous systems are directly referenced within the Act as areas subject to heightened governance obligations [4]. Specifically, employment and creditworthiness assessment systems, as well as certain biometric identification systems, are classified as high risk pursuant to Article 6(2) in conjunction with Annex III of the AI Act. By contrast, certain healthcare applications and safety-critical autonomous systems fall within the high-risk category under Article 6(1), where the AI system is intended to be used as a safety component of a product covered by Union harmonization legislation listed in Annex I (such as healthcare applications and autonomous systems). At the same time, real-world deployments in these sectors illustrate how bias-related harms can emerge despite formal regulatory compliance, particularly in relation to dataset governance, fairness evaluation, continuous monitoring, and accountability attribution. This alignment allows for a systematic mapping between observed technical failures and regulatory control points, facilitating a critical assessment of where and why existing governance mechanisms remain insufficient.

Across all sectors, the analysis applies the same analytical framework by mapping bias mechanisms emerging across the AI lifecycle to the regulatory control points established under the EU AI Act, enabling consistent cross-sector comparison. Taken together, the cross-sectoral analysis indicates that AI bias rarely stems from a single isolated source but rather emerges through the interaction of multiple reinforcing drivers. As depicted in Figure 3, technical bias mechanisms related to data and model design intersect with socio-technical amplification processes driven by human reliance and organizational practices while simultaneously operating within regulatory environments that leave critical aspects of bias unaddressed. These overlapping dynamics give rise to feedback driven and institutionalized forms of bias that persist across sectors, even in systems that are formally governed or compliant. The figure highlights how unaddressed technical bias, socio-technical reinforcement, and regulatory gaps mutually reinforce one another, producing systemic outcomes that cannot be adequately explained or mitigated through isolated technical or legal interventions. This convergence underscores the need for governance approaches that explicitly account for the interaction between technical mechanisms, organizational dynamics, and regulatory structures across the full AI lifecycle.

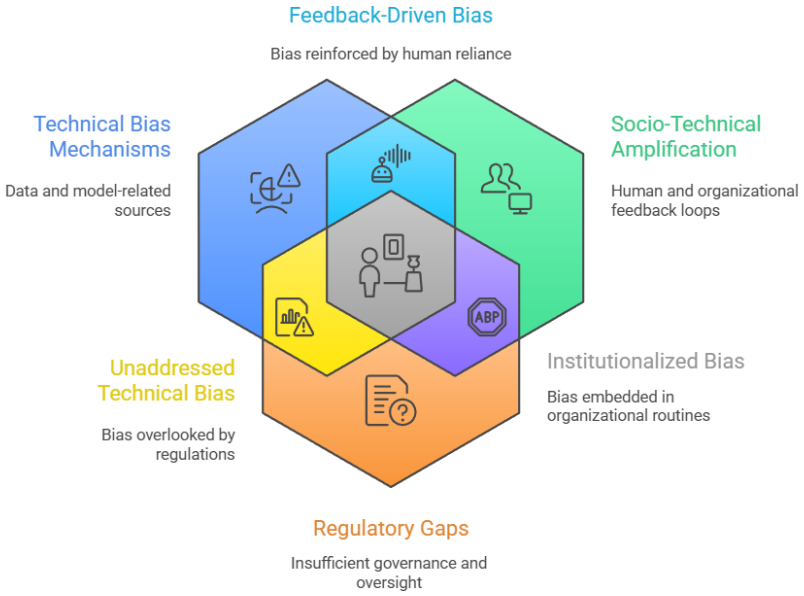


Figure 3. Overlapping drivers of AI bias across sectors, illustrating how technical bias mechanisms, socio-technical amplification, and regulatory gaps interact to produce systemic and persistent bias in real-world AI systems.

4.2. Analytical Dimensions per Sector

For each selected sector, the analysis applies a consistent set of analytical dimensions to ensure comparability across domains. Specifically, each case examines the technical function of the AI system, the sources of data bias across the AI lifecycle, the resulting failure manifestations in deployment, and the corresponding classification and compliance implications under the EU AI Act's risk-based framework. This structured approach enables a systematic identification of recurring bias mechanisms and regulatory blind spots across sectors.

4.2.1. Employment/Hiring

Technical function. AI systems in employment and hiring are commonly deployed to support or automate candidate screening, ranking, and shortlisting processes. These systems typically rely on supervised machine learning models trained on historical hiring data, résumés, psychometric assessments, and performance proxies to predict candidate suitability or job performance [53]. Automated hiring tools are often integrated into applicant tracking systems and operate at scale, filtering large applicant pools before human review. While such systems are framed as efficiency enhancing and decision support tools, their outputs frequently exert a decisive influence on downstream hiring decisions.

Data bias source. Bias in hiring systems primarily originates from historical and representational distortions embedded in training datasets [54]. Historical hiring data often reflect structural inequalities in labor markets, including gender, racial, and socioeconomic disparities, which are implicitly encoded as *successful* patterns. Additional bias arises from proxy variables such as educational background, employment gaps, zip codes, or linguistic features in résumés that correlate with protected characteristics [54]. Labeling practices may further reinforce bias when past hiring decisions are treated as ground truth without critical assessment. As a result, biased assumptions about merit and suitability are introduced early in the data pipeline and propagated through model training.

Failure manifestation. In deployment, biased hiring systems manifest through systematic exclusion or down ranking of certain demographic groups, unequal rejection rates, and reduced visibility of qualified candidates from underrepresented populations. These failures are often difficult to detect due to the opacity of ranking algorithms and the normal-

ization of automated filtering as an objective pre-selection step. Over time, feedback effects may reinforce bias, as the system continues to learn from outcomes shaped by its own prior recommendations. Such failures illustrate how data bias in hiring does not remain a technical artifact but translates into discriminatory outcomes with direct legal and societal implications, frequently without triggering immediate regulatory intervention.

AI Act classification. Under the EU AI Act, AI systems used in employment, workers' management, and access to self employment are explicitly designated as *high risk* pursuant to Article 6(2) in conjunction with Annex III(4). This includes systems intended for the recruitment and selection of natural persons, the analysis and filtering of job applications, the evaluation of candidates, as well as systems used to make decisions affecting terms of employment, promotion or termination, task allocation, and performance monitoring. Recital 57 grounds this classification in the potentially far reaching impact of such systems on individuals' livelihoods, career prospects, and the effective enjoyment of fundamental rights, in particular the rights to non-discrimination, data protection, and privacy. The risk-based framework therefore treats algorithmic hiring not as a neutral productivity tool, but as a category of AI use that is structurally capable of producing legally relevant harm.

Allocation of responsibility: employer as deployer (and occasionally provider). In workplace contexts, the employer is typically the *deployer* within the meaning of Article 3(4), as the entity that uses the AI system under its authority and determines its purpose, scope, and integration into organizational processes. Where the employer also develops and puts the system into service under its own name, it may additionally qualify as a *provider* under Article 3(3), thereby assuming the corresponding conformity assessment obligations, including the internal control procedure under Article 43(2). By contrast, workers who are subject to, or interact with, the system in the course of their employment do not ordinarily constitute deployers, as they lack control over system design, but they are considered as *affected persons that are located in the Union* under Article 2(1)(g) [55].

Compliance implications. As deployers of high risk AI systems, employers are subject to the obligations set out in Article 26. These include ensuring use in accordance with the provider's instructions, assigning effective human oversight, monitoring system operation, retaining logs, and, crucially in the employment context, informing workers' representatives and affected workers prior to workplace deployment (Article 26(7)). Where the employer exercises control over input data, it must also ensure that such data are relevant and sufficiently representative in view of the system's intended purpose (Article 26(4)), directly linking data governance to the risk of discriminatory outcomes identified above. Complementarily, providers must comply with ex ante risk management, dataset quality, documentation, transparency, and post market monitoring duties, thereby embedding bias as a legally cognizable compliance risk within the high risk regime.

Regulatory blind spot. Despite this strict formal classification, the AI Act addresses bias in employment primarily through procedural and technical obligations risk management, documentation and human oversight rather than through outcome oriented equality standards, while Recital 57 explicitly acknowledges the danger of *perpetuating historical patterns of discrimination* when AI is used in employment, the AI Act offers limited guidance on how to detect and remediate structural bias embedded in labor market data or in proxy variables correlated with protected characteristics, while this information obligation strengthens procedural transparency in employment related AI deployments, its protective reach remains limited. The duty is confined to ex ante notification and does not create a substantive right to contest, suspend, or renegotiate the use of high risk AI systems or challenge AI decisions [56], nor does it guarantee effective worker participation [57] in system design, data governance, or impact assessment. Moreover, compliance is largely assessed ex ante at the level of system design and conformity, whereas discriminatory effects often

emerge only in deployment through feedback loops and organizational practices. As a result, hiring systems may satisfy the formal requirements of the high risk regime while continuing to reproduce unequal access to employment. Although Article 26 introduces an information obligation towards workers and their representatives, it notably does not confer any right of access to the underlying the data collected or parameters used [58]. Such access rights do, however, exist in Directive (EU) 2024/2831 on improving working conditions in platform work, which establishes enhanced transparency obligations, including rights for platform workers to receive information regarding automated monitoring and decision-making systems. In particular, Articles 9–11 of that Directive require digital labor platforms to provide detailed information regarding the use of automated monitoring and automated decision-making systems, including the categories of data processed, the main parameters considered and their relative importance, as well as to make relevant operational information available to competent authorities and workers' representatives, plus it further provides rights to explanation, human oversight, review, and rectification of automated decisions. The existence of these strengthened safeguards in the platform economy suggests that, in such algorithmically intensive workplaces, the Union legislature considered broader access rights necessary, protections that are not horizontally embedded in the AI Act's general employment framework. Thus, outside such sector-specific regimes, workers remain unable to receive information on how their personal data, behavioral traces, or inferred attributes are processed, nor can they meaningfully assess whether proxy variables or historically skewed datasets shape outcomes that affect their career progression or their prospects in recruitment and selection processes.

Analytical implication. The employment sector thus exemplifies a central tension of the EU AI Act's risk-based approach: although algorithmic management or hiring is correctly categorized as high risk, the regulatory model remains predominantly process oriented and compliance driven since it treats the AI system as a product that needs to be sold from the provider to the deployer. This creates a potential gap between formal conformity and substantive non-discrimination, revealing a regulatory blind spot in situations where bias is not an isolated technical defect but rather the product of historically structured inequalities encoded in data and organizational decision making. In this sense, workplace AI illustrates how data bias translates into legally significant discrimination without necessarily triggering effective intervention under the current risk-based framework.

4.2.2. Credit and Financial Scoring

Technical function. AI systems in credit and financial scoring are used to assess creditworthiness, determine loan eligibility, set credit limits, and price financial products [59]. These systems typically rely on supervised machine learning or statistical models trained on historical financial data, including repayment histories, transaction records, income proxies, and behavioral indicators [59]. Outputs are often expressed as risk scores or categorical decisions that directly influence access to credit and financial services. Given their integration into automated decision pipelines, such systems frequently operate with limited human intervention and at significant scale.

Data bias source. Bias in credit scoring systems primarily originates from historical financial data that reflects long-standing structural inequalities in access to credit, employment, and wealth accumulation [60]. Protected characteristics are rarely used explicitly, while proxy variables such as postal codes, spending patterns, employment stability, or prior credit history are strongly correlated with socioeconomic status, race, and gender [60]. Measurement bias may also arise when financial behavior is used as a proxy for creditworthiness without accounting for unequal access to financial products or services. As a result,

disadvantaged groups are often underrepresented or systematically encoded as higher risk in training datasets, embedding bias at early stages of the AI lifecycle.

Failure manifestation. In deployment, biased credit scoring systems manifest through systematically lower credit limits, higher interest rates, or outright denial of credit for certain demographic groups, even when financial profiles are comparable. These outcomes are often difficult to contest due to the opacity of scoring models and the normalization of automated risk assessment in financial decision making. Over time, such failures can become self reinforcing, as restricted access to credit negatively affects future financial data, further entrenching disadvantage. These manifestations illustrate how data bias in financial scoring translates into persistent economic exclusion and raise significant legal concerns regarding discrimination, transparency, and accountability.

AI Act classification. Notably, credit scoring sits at the normative boundary with the prohibition of unacceptable AI practices under Article 5(1)(c). According to the Commission Guidelines [61] and the case law of the Court of Justice, most notably *SCHUFA I* [62], AI systems that generate a probability value concerning a person's future behavior on the basis of personal characteristics constitute *profiling* and may amount to social scoring where they lead to detrimental or unfavorable treatment based on unrelated or illegitimate factors. Recital 31 defines social scoring as the evaluation or classification of natural persons based on multiple data points related to their social behavior across contexts or on known, inferred, or predicted personal or personality characteristics over time, where the resulting score leads to detrimental or unfavorable treatment in contexts unrelated to the original data collection or to treatment that is unjustified or disproportionate. Article 5(1)(c) accordingly prohibits the placing on the market, putting into service, or use of AI systems that generate such scores where they result in detrimental or unfavorable treatment of certain natural persons or groups of persons: (i) in social contexts that are unrelated to the contexts in which the data were originally generated or collected and/or (ii) are unjustified or disproportionate to their social behavior or its gravity. Importantly, Article 5(1)(c) does not require the AI generated score to be the sole cause of the adverse outcome, nor does it confine the prohibition to public authorities: harmful scoring practices may fall within the scope of the ban even where the score is produced by one entity and used by another, including in the private sector [63]. Accordingly, when credit scoring relies on personal characteristics or behavioral inferences that are not relevant to legitimate financial assessment, it may move from the high risk regime into the category of prohibited AI.

Where credit scoring is not prohibited, AI systems intended to evaluate the creditworthiness of natural persons or to establish their credit score are expressly designated as *high risk* pursuant to Article 6(2) in conjunction with Annex III(5)(b). Recital 58 grounds this classification in the decisive impact of such systems on access to essential private services and financial resources, including housing, utilities, and telecommunications, and in their documented propensity to perpetuate historical patterns of discrimination. By contrast, AI systems used for fraud detection (Annex III(5)(b)) or for prudential purposes in calculating credit institutions' and insurance undertakings' capital requirements (Recital 58) are explicitly excluded from the high risk category.

Furthermore, financial credit scoring remains outside Article 5(1)(c), where it is based on information relevant and proportionate to the legitimate purpose of assessing creditworthiness and complies with applicable consumer protection law. In particular, Article 18(3) of the Consumer Credit Directive (EU) 2023/2225 [64] requires that assessments be grounded in accurate information on income, expenses, and other financial and economic circumstances; prohibits the use of special categories of personal data and the sourcing of data from social networks; and is complemented by the European Banking Authority's Guidelines [65] on loan origination and monitoring. These sectoral instruments therefore define which

data may lawfully be used for creditworthiness assessments and serve as a benchmark for distinguishing legitimate financial evaluation from prohibited social scoring.

Regulatory blind spot. Notwithstanding this stringent formal architecture, the AI Act's approach to credit scoring remains predominantly procedural. First, the boundary between lawful high-risk scoring and prohibited social scoring hinges on the relevance and proportionality of the data used, yet offers limited operational criteria for assessing when proxies or behavioral inferences become *unrelated* to legitimate financial purposes. Second, while Article 10 embeds bias detection within data governance, it does not ensure that affected individuals can access or scrutinize the features, assumptions, or parameters that shape their scores, thereby constraining effective contestation of discriminatory outcomes. Third, compliance is largely evaluated ex ante at the level of system design and conformity, whereas the most pernicious effects of biased scoring often materialize ex post through cumulative disadvantage and feedback loops in financial histories. Consequently, credit scoring systems may satisfy the formal requirements of the high risk regime while continuing to produce persistent patterns of economic exclusion.

Analytical implication. Credit and financial scoring thus exemplify the limits of the AI Act's risk-based framework; although such systems are correctly classified as high risk and subject to robust data governance duties, the regulation struggles to police the substantive line between legitimate financial assessment and socially harmful profiling and to translate procedural compliance into material non-discrimination. This sector, therefore, exposes a regulatory blind spot analogous to that observed in employment; bias rooted in data infrastructures and proxy-based modeling can remain legally insulated despite the presence of detailed governance obligations.

4.2.3. Healthcare Risk Prediction

Technical function. AI systems for healthcare risk prediction are deployed to estimate patient risk, prioritize care, allocate medical resources, and support clinical decision-making [66]. These systems commonly use supervised machine learning models trained on electronic health records, insurance claims, diagnostic codes, and treatment histories to predict outcomes such as disease progression, hospitalization risk, or healthcare utilization. Outputs are typically expressed as risk scores or stratifications that influence clinical pathways, care management programs, and access to medical interventions; while often positioned as decision support tools, these systems can exert substantial influence over downstream clinical decisions [66].

Data bias source. Bias in healthcare risk prediction systems primarily arises from the use of incomplete or distorted health data that reflect unequal access to care and systemic disparities in healthcare delivery [67]. Training datasets may under-represent certain populations or encode historical differences in diagnosis rates, treatment intensity, and healthcare utilization. Proxy variables, such as healthcare expenditure or frequency of medical visits, are frequently used as substitutes for underlying health needs, introducing measurement bias when cost or utilization does not accurately reflect clinical severity. Additionally, labeling practices based on prior clinical decisions may reproduce existing biases in diagnosis and treatment. These factors introduce bias at early stages of data collection and annotation, which is subsequently propagated through model training [67].

Failure manifestation. In deployment, biased healthcare risk prediction systems manifest through the systematic underestimation of medical risk for certain demographic groups, leading to reduced prioritization for care, delayed interventions, or exclusion from targeted health programs. Such failures may remain invisible within aggregate performance metrics while producing substantial disparities at the subgroup level. Over time, these biased outcomes can reinforce existing health inequities by shaping future data and clinical

practices. These manifestations illustrate how data bias in healthcare AI systems can translate into tangible harm, raising significant ethical, legal, and governance concerns regarding fairness, accountability, and patient safety.

AI Act classification. Healthcare AI systems used for risk prediction are covered by the EU AI Act only where they fall within the category of high risk AI systems linked to regulated medical devices. Recital 12 adopts a broad concept of AI systems based on varying degrees of autonomy and independence from human intervention; thus, a wide range of medical software is considered an AI system [68]. However, Article 6(1) limits the material scope to AI systems that are medical devices or safety components themselves or those that undergo third party conformity assessment covered under the high risk regime by the Union harmonization legislation under Annex I No 11 & 12; thereof only those falling under the Medical Devices Regulation (MDR) [69] or In Vitro Diagnostics Regulation (IVDR) [70] (e.g., AI Clinical Decision Support Systems). Recital 51 reinforces that the MDR/IVDR classification of the device determines whether the AI system is high risk. In practice, this means that most AI driven medical tools and diagnostic algorithms, which typically fall under at least Class IIa medical devices requiring a notified body's review, automatically qualify as *high risk* AI systems in regard to the EU AI Act. Similarly, where an AI system lacks a medical intended purpose set out in the definition of Article 2(1) of the MDR and therefore does not qualify as a medical device (e.g., a lifestyle or administrative application) or is a low risk Class I device, it is generally classified as minimal risk under the AI Act and falls outside the high risk regulatory framework [71]. For example, an AI that predicts hospital readmission rates for internal capacity planning might qualify as medical device software under MDR's broad definition yet remain Class I (no therapeutic/diagnostic purpose); hence, outside the AI Act's high risk scope. The Act also carves out certain in house AI medical devices developed and used within healthcare institutions. If a tool is not *placed on the market* and thus exempt from conformity assessment under MDR, it is not considered high risk AI under Article 6(1)(b). This exclusion for in house tools creates ambiguity; Article 43(3) suggests some of these could still be treated as high risk in the future, implying the need for further guidance from the EC [71]. Once a healthcare AI system falls into the high risk category, it becomes subject to a dual regulatory regime under both the MDR/IVDR and the AI Act. The AI Act does not displace the MDR's safety and performance requirements; rather, it adds an extra layer of AI specific obligations for providers (manufacturers) and deployers (users) of the system. Notified bodies overseeing the device's conformity assessment must integrate the AI Act's requirements into their evaluation (AI Act Articles 16(f) and 43(3)) [68].

Notably, the AI Act extends regulatory responsibilities to healthcare AI users (deployers) in ways that MDR did not. Under Article 26, hospitals and health providers using a high risk AI system must ensure proper oversight and quality control in deployment. This includes maintaining detailed log records for at least six months (or longer, if required by other law) to enable traceability and audit of the AI's operation (AI Act Article 26(6)). Deployers are obligated to implement effective human oversight measures (Article 26(2) (5)), such as workflows for clinicians to review and, if needed, override AI driven risk predictions that seem clinically inappropriate. They must also verify that input data fed into the AI (e.g., patient data used for risk calculations) meets the quality and completeness requirements specified by the provider (Article 26(4)). As Djeffal et al. observe, this represents a shift from traditional tech regulation focused only on manufacturers to a wider range of actors in the AI ecosystem, including distributors and professional users who must now actively ensure ongoing compliance [72].

Regulatory blind spots. Legal scholars have flagged numerous frictions at the intersection of the AI Act and medical device law. One major concern is the lack of a clear

hierarchy or integration between the two regimes, leading to potential inconsistencies [68] in terms of the terminology and taxonomy of AI [73]. For instance, the AI Act defines roles like *provider* and *deployer* differently from MDR's *manufacturer* and *user*, and it classifies risk on a different binary scale (high risk vs. not high risk) rather than MDR's multi class system. This can result in an AI based tool being deemed *high risk* under the AI Act while simultaneously being considered only a moderate risk Class IIa device under MDR, a confusing discrepancy for compliance and oversight. Moreover, although both frameworks impose risk management obligations, they articulate different normative thresholds: risk reduction *as far as possible* under the MDR versus risk minimization subject to proportionality and balance under the AI Act [68]. It is unclear which AI systems operating at the level of health systems or populations, such as tools used for hospital resource allocation, patient recruitment in clinical trials, or the planning and staffing of healthcare services, often do not qualify as medical devices and therefore escape classification as high risk AI. This exclusion extends to many AI applications used in the medicines development cycle, including systems deployed in drug discovery, clinical trials, and pharmaceutical research and promotion strategies, despite their capacity to shape access to treatment and the distribution of healthcare resources. As a result, AI systems with substantial effects on population health may remain unregulated under the AI Act, not because they are low risk in substance, but because they fall outside the narrow boundaries of medical device law. Further, low risk health related systems are subject to minimal oversight, creating a regulatory vacuum that may facilitate the diffusion of ineffective, unproven, or harmful tools and erode public trust in medical AI [74]. Moreover, there is no clear standard that an actor should follow when striving to comply simultaneously with both regulatory regimes. The AI Act does attempt to coordinate with sectoral laws (for example, Article 8(2) allows combining documentation to satisfy multiple regulations), and the expectation is that one conformity assessment can cover both sets of requirements. However, in practice, developers face practical challenges; misaligned timelines, inconsistent terminology, and different responsible parties across MDR, the AI Act, and even GDPR (for data related assessments) make unified compliance difficult [72].

Also, the MDR's own limitation—applying only to products with a medical intended purpose as claimed by the maker—thus carries over into the AI Act, potentially allowing high risk uses of AI (in a commonsense view) to avoid the *high risk* label in law. Because the MDR applies only to products with a medical intended purpose as declared by the manufacturer, this limitation is transposed into the AI Act and allows functionally high impact uses of AI to remain formally outside the high risk regime. As *van Oirschot and Ooms* observe [75], AI systems operating at the level of health systems or populations, such as tools used for hospital resource allocation, patient recruitment in clinical trials, or the planning and staffing of healthcare services, often do not qualify as medical devices and therefore escape classification as high-risk AI. This exclusion extends to many AI applications used in the medicines development cycle, including systems deployed in drug discovery, clinical trials, and pharmaceutical research and promotion strategies, despite their capacity to shape access to treatment and the distribution of healthcare resources. As a result, AI systems with substantial effects on population health may remain unregulated under the AI Act, not because they are low-risk in substance, but because they fall outside the narrow boundaries of medical device law. Moreover, although the AI Act introduces a fundamental rights impact assessment, its procedural design is ambiguous, since it requires identification of potential impacts without imposing a substantive obligation to assess their acceptability or to prevent foreseeable harm [74].

4.2.4. Transport and Autonomous Systems

Technical function. AI systems in transport and autonomous systems are deployed to perform perception, prediction, and control functions in safety critical environments [76]. At the perception layer, these systems rely on deep learning-based computer vision models, including convolutional neural networks and transformer based architectures, for object detection, semantic and instance segmentation, lane detection, and traffic signal recognition, often combined with LiDAR and radar data through sensor fusion techniques [76]. At the prediction layer, sequence and trajectory prediction models estimate the future behavior of dynamic agents such as pedestrians, cyclists, and other vehicles. At the control and planning layer, reinforcement learning, model predictive control, or hybrid rule based approaches are used to translate perceptual inputs into navigation and actuation decisions. Training data typically consist of large scale multimodal datasets combining images, video streams, point clouds, and telemetry collected from real world driving environments, while model outputs directly influence vehicle behavior or traffic management actions. Given their real time operation, closed loop feedback, and limited tolerance for error, these systems are embedded in tightly coupled automated pipelines where perception failures can rapidly propagate into safety critical outcomes [76].

Data bias source. Bias in transport and autonomous systems primarily originates from non-representative training data that insufficiently captures the diversity of real-world conditions. Datasets may over-represent specific environments, such as urban settings, daylight conditions, favorable weather, or particular geographic regions, while under-representing others [76]. Additional bias can arise from disparities in sensor performance, annotation practices, or object labeling, particularly with respect to pedestrians, cyclists, or road users with varying physical characteristics or mobility aids. These limitations introduce systematic distortions at the data collection and labeling stages, which are subsequently reinforced during model training and evaluation [77].

Failure manifestation. In deployment, biased autonomous systems manifest through uneven detection accuracy, delayed responses, or increased error rates under specific conditions or for certain categories of road users. Such failures may disproportionately affect vulnerable populations or occur in less represented environments, leading to elevated safety risks [78]. Because these systems operate at scale and often without immediate human intervention, biased performance can result in repeated and potentially catastrophic outcomes. These manifestations demonstrate how data bias in autonomous systems extends beyond abstract fairness concerns and translates into concrete risks to safety, underscoring the regulatory and governance challenges associated with deploying AI in safety-critical transport contexts. Beyond safety and performance concerns, autonomous transport systems also expose a structural accountability gap, as the diffusion of decision making across designers, manufacturers, and users complicates the attribution of responsibility for harm, while the system itself cannot constitute a bearer of liability [79].

AI Act classification. Under the EU's AI Act, AI systems used in autonomous or driver assistance functions are generally classified as high-risk if they play a safety-critical role in vehicles. In particular, Annex I Section B of the Act lists EU vehicle safety laws, notably the Type Approval Framework Regulation (TAFR) and the General Safety Regulation (GSR) [80], such that any AI system functioning as a *safety component* of a product regulated by those laws is deemed high risk. In practice, the AI-based driver assistance, perception, and decision modules mandated by GSR, such as intelligent speed assistance, lane keeping systems, driver drowsiness detectors, and other advanced driving aids, exemplify these safety components. They rely on AI to interpret sensor input (e.g., recognizing road signals or driver alertness) and to make or assist driving decisions and, thus, fit squarely within the Act's high-risk classification for vehicle safety systems. At the same time, the

distinction between *safety components* and other functional modules is particularly blurred in autonomous vehicles, since virtually all operational layers simultaneously fulfill safety functions and are essential for the overall functioning of the system, making it difficult to delineate which AI components fall within the legal notion of safety components under the Act [81]. Indeed, AI systems deployed on or in connection with *Autonomous Vehicles* (AVs) affecting driving and passenger safety will be considered high risk by default. This encompasses the multiple technological layers from perception (environment sensing and object recognition) and decision/planning (path planning, speed control) to actuation (steering/braking control), each of which is typically powered by AI algorithms [82].

Article 2(2) Carve Out and Sectoral Regime: A crucial legal boundary, however, is that Article 2(2) of the AI Act carves out these very systems from most of the Act's direct requirements. In other words, although classified as high risk, AI systems in the safety components of vehicles are placed under a special regime. This reflects a deliberate choice to avoid frictions with existing AV safety law [83]. Thus, the AI Act has no direct impact on AVs insofar as core vehicle safety is already governed by sectoral regulations, even if these products and systems qualify as high risk AI (Article 2(2)(f)) [84]. Article 2(2) of the Act provides that for high risk AI systems related to products covered by the listed vehicle regulations, only a few provisions of the AI Act apply, notably the classification clause itself (Article 6(1)), certain amendments (Articles 102–109, 112), and a limited obligation regarding regulatory sandboxes if relevant (Article 57, but only to the extent the sectoral law incorporates equivalent requirements). All the substantive obligations of Title III, Chapter 2 of the AI Act (on risk management, data governance, transparency, human oversight, robustness, etc.) do not directly apply to those vehicle AI systems. Instead, the Act mandates a bridging approach. EU motor vehicle legislation must take account of the AI Act's requirements in future implementing or delegated acts for vehicle safety (Article 107 AI Act amending Article 5 of TAFR and Article 109 AI Act amending Article 11 of GSR). *Vellinga* highlights the effect of this carve out: an automated vehicle's AI, clearly high risk by nature, *would not have to be in conformity with the proposed AI Act and its requirements on transparency, robustness, and so on* [85], as long as the vehicle falls under the EU Type Approval or General Safety Regulation regimes. This special treatment creates a lex specialis hierarchy in which vehicle safety AI is overseen primarily through automotive law (with its own conformity assessment and enforcement structure), rather than the AI Act's horizontal framework. In practice, the vehicle regulations in a future act will have to ensure that future rules for vehicle AI systems will *take into account* the mandatory requirements for high risk AI (e.g., on risk management, data governance, and transparency) in a manner adapted to the automotive sector. For AI systems deployed in vehicles that are not covered by the vehicle safety acquis and do not qualify as safety components under Article 6 and Annex I Section B, classification follows the general risk framework of the Act. Scholars point out that the Act's sectoral carve outs leave gaps where potentially risky automotive AI systems are governed only by general product liability rules [85], not by the AI Act's ex ante controls.

4.2.5. Biometric Identification (Facial Recognition)

Technical function. AI systems for biometric identification and, in particular, facial recognition systems are designed to detect, extract, and compare biometric features from facial images or video streams in order to verify or identify individuals. At the perception stage, these systems rely on deep learning-based face detection and alignment models to locate facial regions under varying conditions [86]. Feature extraction is typically performed using deep neural networks that generate high dimensional facial embeddings, which are then compared against reference databases using similarity metrics. Depending on

the application, facial recognition systems may operate in verification mode (one to one matching) or identification mode (one to many matching), with outputs directly influencing access control, surveillance decisions, or identity verification processes. These systems are often deployed in real time or near real time settings and integrated into broader security or administrative infrastructures [86].

Data bias source. Bias in facial recognition systems primarily originates from imbalances and distortions in training datasets. Face datasets frequently over-represent certain demographic groups, particularly individuals with lighter skin tones, specific age ranges, or particular gender expressions, while under-representing others [87]. Variations in image quality, lighting conditions, camera angles, and sensor characteristics further contribute to uneven data distributions across groups. Annotation practices may also introduce bias when demographic attributes are inferred inaccurately or inconsistently. These factors result in feature representations that are optimized for majority groups, embedding bias during the data collection, labeling, and training stages of the AI lifecycle [87].

Failure manifestation. In deployment, biased facial recognition systems manifest through systematically higher error rates for specific demographic groups, including false positives in identification scenarios and false negatives in verification tasks. Such failures can lead to unequal treatment, misidentification, or exclusion, particularly in contexts involving law enforcement, border control, or access to essential services. Because these systems are often perceived as objective and authoritative, their errors may be amplified through institutional reliance and limited avenues for contestation. These manifestations illustrate how data bias in biometric identification systems translates into concrete risks to fundamental rights, accountability, and public trust, raising significant regulatory and governance challenges.

AI Act classification. The AI Act establishes a tiered regulatory framework for biometric systems, combining outright prohibitions with a high-risk classification regime. Article 5(1)(g) prohibits the placing on the market, the putting into service for this specific purpose, or the use of biometric categorization systems that categorize individually natural persons on the basis of biometric data in order to deduce or infer race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation, while expressly excluding the lawful labeling or filtering of biometric datasets and categorization activities in the area of law enforcement from this ban (Article 5(1)(g); Recital 30). Article 5(1)(h) further prohibits the use of *real time* remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, save where such use is strictly necessary for one of three exhaustively listed objectives, namely the targeted search for specific victims or missing persons, the prevention of a specific, substantial and imminent threat to life or physical safety or a genuine terrorist threat, or the localization or identification of a suspect in relation to offences referred to in Annex II punishable by a custodial sentence or detention order of at least four years (Article 5(1)(h)(i)–(iii); Recitals 32–33). Where an exception applies, deployment is limited to confirming the identity of a specifically targeted individual and must take into account the nature of the situation and the consequences for the rights and freedoms of all persons concerned, while complying with necessary and proportionate safeguards, including temporal, geographic and personal limitations, completion of a fundamental rights impact assessment under Article 27 and registration in the EU database under Article 49, subject only to a narrowly framed urgency derogation (Article 5(2); Recital 34). Each use is subject to prior authorization by a judicial or independent administrative authority, with an urgency mechanism requiring a request within 24 h and immediate cessation and deletion of all data and outputs if authorization is refused, while no adverse legal decision may be taken solely on the basis of the system's output (Article 5(3); Recital 35).

Where biometric systems are not prohibited, Article 6(2) in conjunction with Annex III classifies as high-risk biometric AI systems insofar as their use is permitted under relevant Union or national law. Annex III includes high-risk remote biometric identification systems, excluding only AI systems intended solely for biometric verification to confirm that a specific natural person is who they claim to be (Annex III(1)(a); Recital 54). It further classifies as high-risk AI systems intended for biometric categorization according to sensitive or protected attributes and AI systems intended for emotion recognition (Annex III(1)(b)–(c); Recital 54). Recital 54 grounds this classification in the special sensitivity of biometric data and the heightened risk of biased and discriminatory outcomes, particularly in relation to age, ethnicity, race, sex, and disabilities, while expressly excluding from the high-risk category biometric systems used solely for cybersecurity and personal data protection purposes.

As *Kindt* observes [88], biometric AI systems that do not fall under the prohibitions of Article 5 will, in most cases, nevertheless qualify as high risk under Article 6 and Annex III. In particular, an emotion recognition system not used in the workplace or education will not be prohibited, but, insofar as it is intended for emotion recognition, it will fall within the high-risk category pursuant to Annex III(1)(c) and Recital 54. High-risk biometric systems are therefore subject to the full compliance architecture of Title III, including continuous risk management, data governance, accuracy, robustness, and cybersecurity, technical documentation, record keeping, transparency, and human oversight obligations, as well as specific duties for providers and deployers (Articles 9–29). By contrast, biometric verification systems excluded under Annex III(1)(a) remain governed primarily by the GDPR, while biometric systems that are neither prohibited nor classified as high risk fall under the transparency regime of Article 50(3), requiring that deployers of an emotion recognition system or a biometric categorization system shall inform the natural persons exposed thereto of the operation of the system.

Untargeted scraping of facial images. Article 5(1)(e) prohibits the placing on the market, the putting into service for this specific purpose, or the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage. The prohibition applies cumulatively where (i) an AI system is placed on the market, put into service or used, (ii) for the purpose of creating or expanding a facial recognition database, (iii) through untargeted scraping, and (iv) from the internet or CCTV sources. Recital 43 grounds this prohibition in the feeling of mass surveillance and the risk of gross violations of fundamental rights, including the right to privacy. The notion of *untargeted* covers indiscriminate harvesting of facial images without focusing on specific individuals, while targeted searches remain outside the prohibition. The rule applies irrespective of the public availability of images and reflects the absence of any lawful basis for such large scale data collection under Union data protection law [61].

Emotion recognition in workplace and education. Article 5(1)(f) prohibits the placing on the market, putting into service, or use of AI systems to infer emotions of a natural person in the areas of workplace and educational institutions, except where the use is intended for medical or safety reasons. The prohibition covers systems identifying or inferring emotions or intentions on the basis of biometric data (Article 3(39); Recital 44) and applies broadly to recruitment, monitoring, assessment, and learning environments, reflecting the structural imbalance of power in those contexts. Physical states such as pain or fatigue are excluded (Recital 18). Uses falling outside the prohibition, including emotion recognition in other domains, are nevertheless classified as high risk pursuant to Annex III(1)(c) and subject to the full Title III compliance regime, together with the transparency obligations of Article 50(3) [61].

5. Cross Sectoral Findings and Governance Gaps

This section synthesizes the sectoral analyses presented in Section 4 to identify recurring patterns of systemic data bias and the governance gaps that enable them to persist across domains. Moving beyond application specific failures, it examines how shared technical mechanisms interact with the risk-based architecture of the EU AI Act. The analysis highlights a structural misalignment between lifecycle based bias dynamics and post hoc, category driven regulatory interventions.

5.1. Recurring Technical Patterns of Data Bias

As addressed in Section 4.2, the sectoral cases reveal a consistent set of technical bias patterns that recur across all examined domains, despite differences in application context, data modality, and model architecture. These patterns indicate that data bias in real world AI systems is not primarily sector specific but pipeline specific, arising from shared design and data practices embedded throughout the AI lifecycle. Specifically, across all examined sectors, a first recurring technical pattern is the systematic introduction of bias at the first and earliest stages of the AI pipeline, particularly during data collection and labeling. Training datasets are commonly derived from historical records that reflect prior decisions, institutional practices, and structural inequalities but are treated as neutral representations of reality. Labeling processes often reinforce these distortions when past outcomes are used as ground truth without critical assessment. As a result, bias is embedded before model training begins, limiting the effectiveness of downstream technical interventions.

A second recurring pattern concerns the pervasive use of proxy variables to approximate complex or unobservable attributes. Variables such as postal codes, healthcare costs, employment histories, transaction patterns, or behavioral indicators are routinely employed as substitutes for creditworthiness, health need, job suitability, or risk. Although technically convenient and often legally permissible, these proxies are strongly correlated with protected characteristics and social disadvantage. Importantly, the same proxy based logic recurs across domains with different data types and objectives, indicating a domain agnostic bias mechanism rooted in pipeline design rather than sector specific misuse.

A third pattern relates to evaluation practices that prioritize aggregate performance metrics over subgroup level analysis. Across domains, models are typically validated using accuracy or overall error rates that obscure differential performance across populations. This misalignment allows biased systems to pass testing and validation stages, with disparities only becoming visible after deployment, when systems interact with diverse real world users. Finally, across all sectors, bias predominantly manifests during deployment rather than within controlled testing environments. Once operationalized, AI systems shape institutional workflows and data generation processes, creating feedback loops that reinforce initial distortions. These production stage effects are rarely anticipated at design time and are difficult to correct retroactively, particularly when systems operate at scale and with limited human oversight.

Taken together, these recurring technical patterns demonstrate that data bias is a lifecycle-wide phenomenon driven by shared pipeline characteristics rather than by sector-specific idiosyncrasies. Bias is introduced early, propagated through proxy-based modeling and inadequate evaluation, and stabilized through deployment and feedback effects. Recognizing the domain agnostic nature of these patterns is essential for understanding why isolated technical fixes and sector-specific interventions have limited effectiveness in preventing recurring AI system failures.

5.2. Cross Sectoral Synthesis of Bias Mechanisms and Regulatory Gaps

Table 2 consolidates the findings of the sectoral analyses presented in Section 4 into a unified cross-sectoral perspective on systemic data bias and its governance failures under the EU AI Act. Rather than summarizing individual cases, the table extracts recurring bias mechanisms and regulatory gaps that manifest consistently across application domains, irrespective of data modality, model architecture, or sector-specific legal regimes. This synthesis reveals that bias in real-world AI systems follows a lifecycle-driven logic, while regulatory oversight remains predominantly static, procedural, and category-dependent. The analysis identifies three recurrent forms of classification failure within the AI Act's risk-based framework: boundary ambiguity between regulatory categories, scope leakage where materially impactful systems fall outside high-risk classification, and carve-outs or lex specialis regimes that displace horizontal governance obligations.

Table 2. Cross-sectoral synthesis including risk classification gaps under the EU AI Act.

Analytical Dimension	Cross Sectoral Finding	Manifestation Across Sectors	EU AI Act Governance Gap
Bias Origin	Bias introduced early in lifecycle (collection/annotation)	Historical records (hiring/credit), distorted healthcare utilization, imbalanced vision datasets (biometrics/AV)	Art. 10 is largely procedural; no independent/outcome oriented dataset audit requirement
Proxy Variables	Proxies encode protected attributes & structural disadvantage	Zip code (credit), cost/utilization (health), education gaps (hiring), contextual visual cues (AV)	No explicit ex ante constraints or justification duty for proxy use; indirect discrimination handled abstractly
Labeling Provenance	Labels inherit prior human decisions as "ground truth"	Hiring outcomes, clinical decisions, security/policing labels	No requirement to assess decision provenance or embedded bias in labeling practices
Evaluation Regime	Aggregate metrics mask subgroup disparities	Sector wide: overall accuracy hides unequal FPR/FNR across groups	No mandatory fairness metrics/subgroup evaluation under Arts. 9 & 15
Deployment Dynamics	Bias manifests most strongly post deployment	Discriminatory allocation (jobs/credit/care), real world error inflation (biometrics/AV)	Art. 61 monitoring is reactive; no continuous bias reassessment requirement
Feedback Loops	Outputs reshape future data, stabilizing bias	Credit histories, hiring pools, policing data, care pathways	Weak lifecycle linkage between early stage bias and downstream accountability triggers
Human Oversight	Oversight exists formally, weak operationally (automation bias)	Clinicians/HR over rely on AI; security operators trust outputs	Art. 14 lacks operational criteria and measurable intervention thresholds
Transparency & Contestability	Individuals face high evidentiary barriers to challenge outcomes	Opaque scoring/ranking/risk outputs across sectors	Arts. 13, 85–86: limited enforceability; practical access to evidence deficit
Accountability Allocation	Responsibility fragmented across provider–deployer chain	Split control of data, model, deployment settings	No robust mechanism tying bias emergence to enforceable liability attribution
Risk Classification Gap (Boundary & Scope Leakage)	Legal classification fails to track bias propagation and real world harm	(i) Credit scoring sits near Art. 5 social scoring boundary; proxies can shift systems across regimes without clear technical test. (ii) Healthcare tools outside MDR/IVDR escape high risk despite material impacts. (iii) Biometric prohibitions/permissions depend on context; similar tech yields different obligations.	Risk triggers are use context and legal category dependent, not lifecycle bias dependent; weak operational criteria for when systems "move" between prohibited/high risk/minimal risk
Risk Classification Gap (Carve outs/Lex specialis)	High risk in substance, but exempt in practice through sectoral carve outs	Vehicle/AV AI classified high risk yet largely excluded via Art. 2(2) product regime routing	"High risk" label does not guarantee application of Title III obligations; governance displaced to sectoral regimes with non equivalent bias controls
Risk Classification Gap (Static vs. Dynamic Risk)	Risk is treated as static at placing on market, but bias is dynamic	Drift + feedback loops change risk profile after deployment in all sectors	No strong mechanism for reclassification or escalation when bias emerges post deployment

A first cross-sectoral finding concerns the temporal misalignment between bias generation and regulatory intervention. Across all examined sectors, bias is systematically introduced at early stages of the AI lifecycle, particularly during data collection, annotation, and problem formulation, yet regulatory obligations are primarily triggered at the point of market placement or deployment. As a result, historically embedded distortions, proxy-based representations, and biased labeling practices are often treated as compliant inputs, even though they function as structural determinants of downstream discriminatory outcomes. Prior scholarship has warned that AI systems may appear neutral while reproducing entrenched inequalities, noting that machines can *entrench pre-existing inequalities while disguised as cost-effective innovation*, particularly where historical and institutional bias is embedded in the data used to train them [89]. The AI Act's data governance provisions, most notably Article 10, emphasize documentation and procedural safeguards but do not impose outcome-oriented or independently verifiable requirements capable of addressing structural bias.

Second, the synthesis highlights the domain-agnostic role of proxy variables as a dominant bias mechanism. Across employment, credit scoring, healthcare, biometric identification, and autonomous systems, technically convenient proxies are routinely used to approximate complex social or behavioral attributes [90]. Although formally neutral, these proxies are strongly correlated with protected characteristics and social disadvantage, enabling indirect discrimination to persist without explicit reference to sensitive attributes. The table shows that this mechanism recurs across sectors yet remains insufficiently constrained by the regulatory framework, which addresses proxy-based bias only indirectly through abstract non-discrimination principles rather than enforceable ex ante technical obligations.

Third, Table 2 illustrates a consistent evaluation gap across sectors. Model validation practices prioritize aggregate performance metrics that obscure subgroup-level disparities, allowing biased systems to pass conformity assessments and internal testing phases. Discriminatory effects, therefore, tend to materialize primarily after deployment, when systems interact with heterogeneous populations and operate at scale, while the AI Act includes post-market monitoring obligations; these mechanisms are largely reactive and incident-driven, offering limited capacity to detect, reassess, or escalate bias-related risks as they evolve dynamically over time.

In practice, addressing this limitation requires continuous bias monitoring mechanisms capable of detecting disparities as they emerge during system operation. A minimal monitoring loop aligned with the AI Act's post-market monitoring logic would include four elements. First, monitored indicators should track both overall performance and subgroup-level disparities, alongside signals of data drift or shifts in user populations. Second, predefined triggers and thresholds should initiate review when statistically significant performance gaps, repeated complaints, or operational anomalies are detected. Third, monitoring systems should classify bias-related incidents, distinguishing between disparate outcomes affecting protected groups, systematic performance degradation for specific populations, feedback-loop amplification, and failures of logging or transparency that prevent auditability. Fourth, escalation procedures should specify corrective actions and responsibilities, including internal review, model retraining or recalibration, documentation updates, temporary suspension of automated decisions where necessary, and, where relevant, reporting through post-market monitoring channels. Such mechanisms would transform monitoring obligations from a reactive compliance requirement into an operational safeguard against the lifecycle dynamics of systemic bias.

A central contribution of the cross-sectoral synthesis is the identification of a structural classification gap within the EU AI Act's risk-based architecture. As shown in the table,

risk classification is anchored in legal categories, sectoral boundaries, and use context definitions, rather than in the mechanisms through which bias propagates across the AI lifecycle. This results in three recurrent failures: boundary problems at the margins of prohibited and high-risk practices, scope leakage where materially impactful systems fall outside high-risk classification, and carve-outs or lex specialis regimes that exempt substantively high risk systems from horizontal governance obligations. Consequently, systems may be high-risk in substance but not in regulatory coverage, or may remain formally compliant while producing escalating discriminatory outcomes.

Finally, the synthesis underscores a pervasive accountability and contestability deficit. Responsibility for biased outcomes is fragmented across providers, deployers, and institutional users, each exercising partial control over different lifecycle stages. Providers typically control model design, training data selection, and technical documentation; deployers determine the operational context, decision thresholds, and institutional use of outputs; while downstream actors such as importers or distributors may influence system integration or configuration (Figure 4). This distribution of control complicates the attribution of responsibility when bias emerges after deployment, particularly where data governance, model design, and operational use are controlled by different actors. At the same time, affected individuals face high evidentiary and informational barriers when seeking to contest or challenge biased decisions. Transparency and oversight obligations, while formally present, lack the operational specificity required to translate bias detection into enforceable corrective action. Taken together, these findings demonstrate that systemic data bias persists not due to isolated technical failures or sector specific misuse, but because of a fundamental misalignment between lifecycle based bias dynamics and the static, category driven logic of current AI governance.

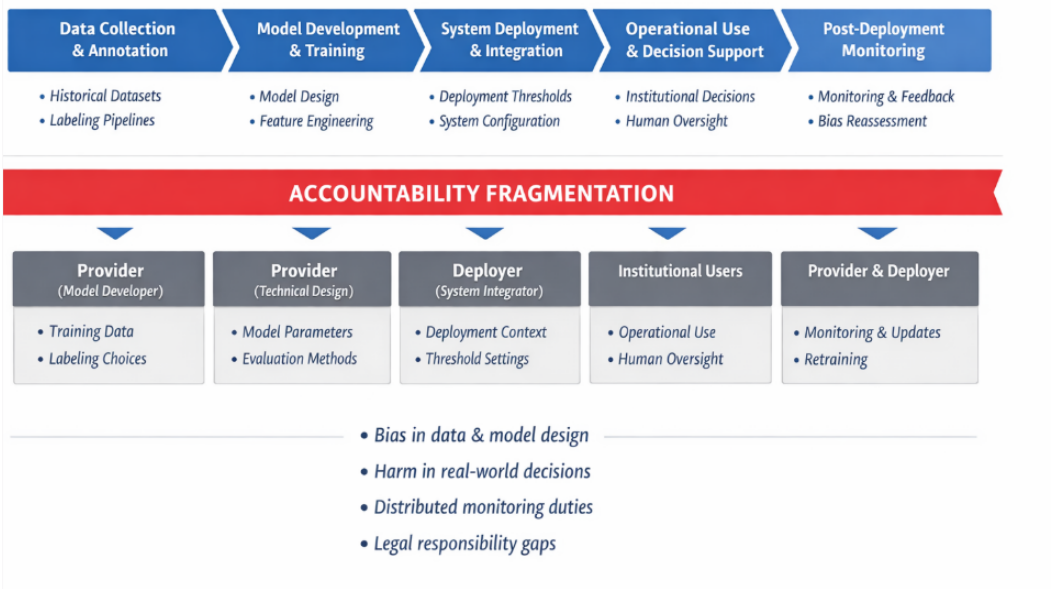


Figure 4. Control over data, model design, deployment configuration, and operational use is distributed across different actors within the AI supply chain. This distribution creates gaps between the locus of technical decision-making and the point at which discriminatory outcomes materialize, complicating accountability under the EU AI Act.

5.3. Implications for Law by Design in AI Governance

The cross-sectoral findings of this study give rise to a limited but significant normative implication for AI governance. If systemic data bias emerges early in the AI lifecycle and propagates through design, deployment, and institutional use, then regulatory interventions that operate exclusively post hoc are structurally incapable of addressing the source

of harm. This observation does not point to the absence of regulation but to a mismatch between the temporal logic of law and the lifecycle dynamics of AI systems. In practice, legal norms already function as design constraints. Requirements related to data governance, risk management, transparency, and human oversight shape how AI systems are specified, trained, and evaluated. However, under the current risk-based architecture of the EU AI Act, these constraints are applied primarily as compliance checks rather than as constitutive elements of system design. As a result, legally relevant values such as non-discrimination and accountability enter the AI lifecycle late, indirectly, and often after bias has been technically stabilized.

From this perspective, law by design should not be understood as an additional regulatory layer or a new governance framework but as an analytical description of how legal norms inevitably interact with technical design choices. Where law is treated as external and post hoc, bias related risks are normalized as technical artifacts. Where legal obligations are internalized as design parameters, bias becomes visible as a legally salient feature of system behavior rather than a residual engineering problem. The relevance of this implication lies not in prescribing specific regulatory reforms but in clarifying the limits of post hoc, category driven governance. As long as legal responsibility remains detached from lifecycle wide design decisions, formal compliance will continue to coexist with systemic discriminatory outcomes. Law by design, in this sense, marks the boundary condition for aligning AI governance with the realities of how biased systems are produced and sustained. In practical governance terms, this perspective implies mechanisms capable of linking legal accountability to technical system design, such as stronger dataset auditability, subgroup-level performance reporting, periodic bias reassessment during deployment, and improved traceability of decision provenance.

Figure 5 depicts the central analytical claim of this paper. Systemic data bias emerges early in the AI lifecycle and propagates through successive technical and institutional stages, while regulatory interventions under the EU AI Act are predominantly triggered after system design choices have already been stabilized. The figure makes explicit the temporal and structural mismatch between lifecycle-wide bias dynamics and post hoc, risk-based governance. In this context, law by design does not function as a policy proposal but as an analytical description of how legal norms necessarily interact with technical design decisions if accountability is to track the production of harm rather than its downstream manifestations.

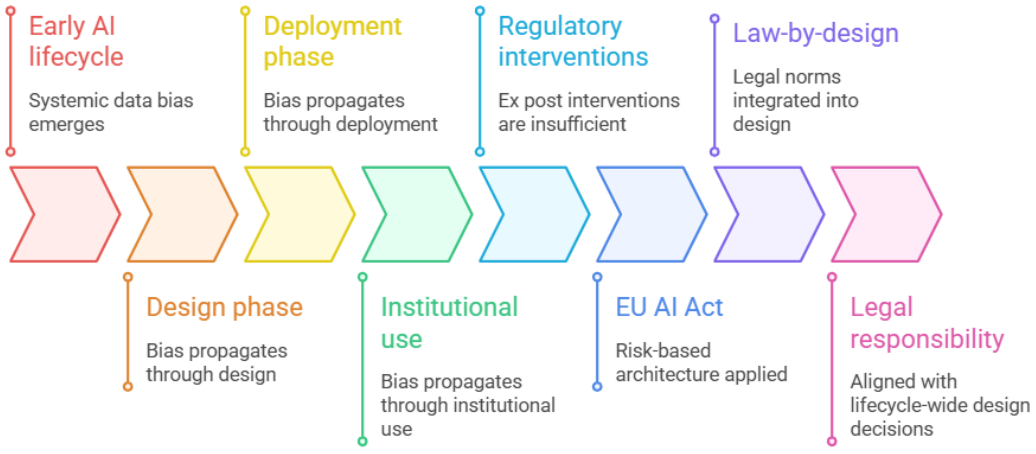


Figure 5. Aligning AI governance with lifecycle dynamics, showing how systemic bias emerges during early design stages while regulatory oversight is typically triggered only after deployment, underscoring the need to integrate legal norms into AI system design.

6. Discussion

6.1. Synthesis of Findings

The cross-sectoral analysis conducted in this study reveals that systemic data bias in real-world AI systems follows recurring lifecycle patterns that transcend individual application domains. Across employment, credit scoring, healthcare, biometric identification, and autonomous systems, bias consistently emerges during early stages of the AI pipeline—particularly in data collection, annotation practices, and problem formulation. These initial distortions are subsequently propagated through proxy-based modeling strategies and evaluation regimes that prioritize aggregate performance metrics over subgroup-level analysis.

A key insight from the sectoral synthesis is that the persistence of bias cannot be attributed solely to isolated technical failures. Instead, bias arises from the interaction between technical design decisions, socio-technical deployment dynamics, and institutional reliance on algorithmic outputs. Once deployed, AI systems often shape organizational practices and data generation processes, creating feedback loops that reinforce initial distortions and stabilize unequal outcomes over time.

At the same time, the analysis demonstrates that the regulatory architecture of the EU AI Act does not fully align with these lifecycle dynamics. Governance interventions are predominantly triggered at the point of market placement or system deployment, whereas the most consequential bias mechanisms often originate earlier in the development process. This temporal misalignment allows structurally biased systems to remain formally compliant while continuing to produce discriminatory outcomes in practice.

6.2. Governance Implications

The findings of this study suggest that addressing systemic bias requires governance approaches that extend beyond post hoc compliance mechanisms and engage more directly with technical design processes across the AI lifecycle. In particular, regulatory frameworks that focus primarily on risk classification at the point of deployment may struggle to address bias mechanisms embedded in data practices, proxy-variable selection, and evaluation regimes.

A lifecycle-oriented governance perspective implies that legal accountability mechanisms must interact more directly with system development practices. This includes greater attention to dataset provenance and representativeness, more transparent justification of proxy variables used in predictive modeling, and evaluation procedures capable of detecting subgroup-level disparities rather than relying solely on aggregate accuracy metrics. In addition, the dynamic nature of AI system behavior after deployment suggests the need for governance mechanisms capable of monitoring emerging bias through continuous reassessment rather than relying exclusively on static conformity assessments.

Importantly, these implications do not necessarily require the introduction of entirely new regulatory regimes. Instead, they highlight the importance of aligning existing legal obligations,-such as data governance, transparency, human oversight, and post-market monitoring, with the technical realities of how bias emerges and propagates within AI systems.

6.3. Lessons Learned

Several broader lessons emerge from the analysis. First, systemic bias in AI systems is fundamentally a lifecycle phenomenon rather than a localized technical error. Interventions that focus narrowly on model optimization or late-stage compliance checks are therefore unlikely to address the underlying drivers of bias. Second, bias mechanisms tend to be domain-agnostic. Despite differences in application context, data modality, and system

architecture, similar technical patterns recur across sectors, particularly in relation to historical training data, proxy-variable use, and evaluation practices. This suggests that effective governance responses should focus on shared pipeline dynamics rather than exclusively sector-specific regulation.

Third, governance gaps often arise not from the absence of regulatory frameworks but from a misalignment between the structure of legal oversight and the technical processes through which AI systems are developed and deployed. The EU AI Act introduces important safeguards but largely treats risk as a static classification problem rather than a dynamic process that evolves throughout the system lifecycle. Finally, addressing systemic bias requires recognizing that legal norms and technical design decisions are deeply interconnected. Treating governance as an external constraint applied after system development risks leaving the most consequential sources of bias unaddressed. By contrast, approaches that integrate legal accountability considerations into system design processes offer a more promising pathway for aligning technological innovation with principles of fairness, accountability, and non-discrimination.

6.4. Future Research

Future research could extend the present analysis through empirical investigation of specific real-world deployments of AI systems. While this study identifies recurring bias mechanisms and governance gaps through cross-sectoral analysis, detailed empirical case studies of operational AI systems could provide deeper insight into how these mechanisms manifest within particular institutional contexts. Such work could examine concrete deployments in domains such as hiring systems, healthcare decision-support tools, financial risk assessment platforms, or biometric identification technologies, allowing researchers to analyze the interaction between system design, data practices, organizational use, and regulatory oversight in practice.

A second direction for future research concerns the rapid emergence of foundation models and large model agents capable of integrating reasoning, planning, and tool use across multiple domains. As these systems become embedded in complex socio-technical environments, they may introduce new governance challenges related to bias propagation, transparency, monitoring, and accountability. Investigating how lifecycle-oriented governance approaches apply to these increasingly general-purpose AI architectures represents an important area for future work.

6.5. Limitations

This study has some limitations that should be acknowledged. First, the analysis does not constitute an empirical audit of datasets or AI systems across sectors. Instead, it relies on documented cases and examples from the technical and policy literature in order to identify recurring bias mechanisms and governance gaps. Second, the sectoral illustrations examined in the paper are intended to highlight structural patterns rather than provide an exhaustive survey of all AI deployments in these domains. Third, the interpretation of the EU Artificial Intelligence Act presented in this analysis reflects the current regulatory framework and available scholarship. As the Act is implemented in practice, further regulatory guidance, delegated acts, and judicial interpretation may refine aspects of the governance architecture discussed here. These limitations do not undermine the central analytical claim of the paper but rather indicate areas where future empirical and doctrinal research could further extend the analysis.

7. Conclusions

This paper examined systemic data bias in real-world AI systems through a cross-sectoral and lifecycle-based analytical lens, focusing on the interaction between technical bias mechanisms and the regulatory architecture of the EU AI Act. Drawing on sectoral analyses across employment, credit scoring, healthcare, biometric identification, and autonomous systems, the study demonstrated that data bias is neither an isolated technical defect nor a sector-specific anomaly. Instead, it emerges early in the AI lifecycle, propagates through design and deployment choices, and stabilizes through institutional use, producing discriminatory outcomes even in formally compliant systems.

A central contribution of the analysis lies in showing that the persistence of bias cannot be adequately explained by gaps in implementation or enforcement alone. Rather, it reflects a structural misalignment between lifecycle-wide bias dynamics and the static, category-driven logic of risk-based AI governance. The EU AI Act, while advancing important safeguards, primarily intervenes at points where critical design choices have already been made, and treats risk as a legally classifiable property rather than as a dynamic process shaped by data practices, proxy use, evaluation regimes, and post-deployment feedback loops. As a result, systems may satisfy regulatory requirements while continuing to generate substantively harmful and unequal outcomes.

By synthesizing sectoral findings into a cross-sectoral framework, the paper identified recurring technical patterns of bias and a structural classification gap within the current regulatory approach. This gap allows systems to be high risk in substance but not in governance coverage or to remain legally acceptable despite escalating bias after deployment. The analysis further highlighted how accountability and contestability deficits compound this problem, fragmenting responsibility across actors and limiting the capacity of affected individuals to challenge biased decisions in practice.

Overall, these findings underscore a fundamental limit of post hoc, compliance oriented AI governance. Addressing systemic data bias requires aligning legal responsibility with the full AI lifecycle and recognizing that legal norms inevitably shape, and are shaped by, technical design decisions. While this paper does not propose a new regulatory framework, it clarifies the conditions under which existing risk-based approaches fall short and delineates the boundary beyond which post hoc regulation alone cannot meaningfully govern biased AI systems. In doing so, it contributes to ongoing debates on AI governance by reframing data bias as a structural coordination problem between technological design and legal oversight, rather than as a residual issue to be corrected after systems are already in use.

In practical terms, the findings of this study also suggest a set of governance practices that could help align technical system development with legal accountability across the AI lifecycle. Rather than constituting a new regulatory framework, these practices summarize operational implications that follow directly from the cross-sectoral analysis of bias mechanisms and governance gaps identified in this paper.

1. Dataset auditability: Maintain traceable documentation of dataset sources, collection procedures, and annotation practices.
2. Label provenance review: Assess whether labels used as ground truth reproduce prior institutional decisions or historical bias.
3. Proxy variable justification: Evaluate and document the use of proxy variables correlated with protected characteristics or structural disadvantage.
4. Subgroup-level evaluation: Complement aggregate performance metrics with subgroup error analysis to identify differential system behavior.
5. Evaluation transparency: Document evaluation methods and metric choices to support regulatory review and external scrutiny.

6. Bias monitoring triggers: Establish predefined indicators and thresholds that trigger reassessment when disparities emerge during deployment.
7. Post-deployment bias reassessment: Periodically evaluate operational systems for drift, feedback loops, and emerging performance disparities.
8. Decision provenance traceability: Maintain logs linking model outputs, decision thresholds, and institutional decision processes.
9. Contestability evidence: Ensure that documentation and system records enable affected individuals or regulators to reconstruct and challenge automated decisions.
10. Lifecycle governance integration: Align compliance mechanisms with the full AI lifecycle rather than relying exclusively on post hoc regulatory intervention.

Author Contributions: Conceptualization, T.F. and A.D.; methodology, T.F. and A.D.; validation, A.D., and T.F.; formal analysis, A.D. and T.F.; investigation, A.D. and T.F.; resources, A.D. and C.-N.A.; writing—original draft preparation, A.D. and T.F.; writing—review and editing, T.F., A.D., and C.-N.A.; visualization, A.D.; supervision, C.-N.A.; project administration, T.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lawrence Livermore National Laboratory. The Birth of Artificial Intelligence (AI) Research | Science and Technology. St.llnl.gov, Science and Technology. Available online: https://st.llnl.gov/news/look-back/birth-artificial-intelligence-ai-research (accessed on 27 December 2025).
2. Verle, L. Deep Blue x Kasparov: A revanche. *Rev. FAMECOS* **1998**, *5*, 63–67. [CrossRef]
3. Jones, M.L. AI in History. *Am. Hist. Rev.* **2023**, *128*, 1360–1367. [CrossRef]
4. The EU Artificial Intelligence Act. Available online: https://artificialintelligenceact.eu/ (accessed on 27 December 2025).
5. Phang, K.; Kaabi, J. Privacy in Flux: A 35-Year Systematic Review of Legal Evolution, Effectiveness, and Global Challenges (US/EU Focus with International Comparisons). *J. Cybersecur. Priv.* **2025**, *5*, 103. [CrossRef]
6. Njoto, S.; Cheong, M.; Frermann, L.; Ruppanner, L. Bias and Discrimination Against Women and Parents in Semi-Automated Hiring Systems. *New Technol. Work. Employ.* **2025**, *40*, 436–446. [CrossRef]
7. Hurlin, C.; Pérignon, C.; Saurin, S. The Fairness of Credit Scoring Models. *arXiv* 2022. [CrossRef]
8. Chua, M.; Kim, D.; Choi, J.; Lee, N.G.; Deshpande, V.; Schwab, J.; Lev, M.H.; Gonzalez, R.G.; Gee, M.S.; Do, S. Tackling prediction uncertainty in machine learning for healthcare. *Nat. Biomed. Eng.* **2023**, *7*, 711–718. [CrossRef] [PubMed]
9. Lai, K.; Yanushkevich, S.N.; Shmerko, V. Fairness, bias and trust in the context of biometric-enabled autonomous decision support. In *Transactions on Computational Science XL*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 66–87.
10. Bose, M. Bias in AI: A societal threat: A look beyond the tech. In *Open AI and Computational Intelligence for Society 5.0*; IGI Global Scientific Publishing: Palmdale, PA, USA, 2025; pp. 197–224.
11. Veale, M.; Borgesius, F.Z. Demystifying the draft EU artificial intelligence act. *arXiv* **2021**. [CrossRef]
12. Falelakis, T. AI-driven tools as democratic equalizers for access to justice. In *AI and the Future of Democracy*; Vajjhala, N.R., Jacob, J.U.-U., Eds.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2025; pp. 168–189. [CrossRef]
13. Fisher, A.E.; Fisher, A.E.; Allday, R.A.; Jones, M.; Samudre, M.D. The impact of a short and explicit labeling bias video on preservice educator behavioral expectations. *J. Educ. Stud. Placed Risk (JESPAR)* **2024**, *29*, 105–129. [CrossRef]
14. Duong Khoi, M.; Conrad, S. Measuring and mitigating bias for tabular datasets with multiple protected attributes. *arXiv* **2024**, arXiv:2405.19300v3. [CrossRef]
15. Zhu, J.; Salimi, B. Overcoming Data Biases: Towards Enhanced Accuracy and Reliability in Machine Learning. *IEEE Data Eng. Bull.* **2024**, *47*, 18–35.

16. Markova, V. Toward Fair and Interpretable AI: A Regulation-Aware Framework for Detecting Bias and Mitigating Unfairness. Master's Thesis, Universidad Politécnica de Madrid, Madrid, Spain, 2025. Available online: https://oa.upm.es/90801/1/TFM_VICTORIA_MARKOVA.pdf (accessed on 20 January 2026).
17. Nishant, R.; Schneckenberg, D.; Ravishankar, M.N. The formal rationality of artificial intelligence-based algorithms and the problem of bias. *J. Inf. Technol.* **2024**, *39*, 19–40. [CrossRef]
18. Alabi, M. Ethical Challenges in AI: Addressing Bias, Privacy, and Accountability. ResearchGate. 2025. Available online: www.researchgate.net/publication/390582838_Ethical_Challenges_in_AI_Addressing_Bias_Privacy_and_Accountability (accessed on 20 January 2026).
19. Peterson, B. Ethical Considerations of AI in Software Engineering: Bias, Reliability, and Human Oversight. ResearchGate. 28 March 2025. Available online: www.researchgate.net/publication/390280753_Ethical_Considerations_of_AI_in_Software_Engineering_Bias_Reliability_and_Human_Oversight (accessed on 20 January 2026).
20. Ravi, K. Socio-Technical System Challenges in the Era of Artificial Intelligence: A Comprehensive Analysis. *Int. J. Bus. Manag. Stud.* **2025**, *6*, 75–91. [CrossRef]
21. Rieskamp, J.; Budnik, Y.; Mirbabaie, M.Why AI Deployment Fails in Organisations: A Socio-technical Perspective on the Root Causes. In Proceedings of the Australasian Conference on Information Systems (ACIS 2025). UniSC AAIS. Available online: https://www.researchgate.net/publication/396929337_Why_AI_Deployment_Fails_in_Organisations_A_Socio-_Technical_Perspective_on_the_Root_Causes (accessed on 20 January 2026).
22. Bruneau, G.A. The Bias Network Approach: A Sociotechnical Approach to Aid AI Developers to Contextualise and Address Biases. Ph.D. Thesis, University of Leeds, Leeds, UK, 2024. Available online: https://etheses.whiterose.ac.uk/id/eprint/36426/ (accessed on 20 January 2026).
23. Smacchia, M.; Za, S.; Arenas, A.E. Identifying AI Bias and Mitigation Challenges Through a Socio-Technical Perspective. *Ecis 2024 Proc.* **2024**, *12*, 1–8. Available online: https://aisel.aisnet.org/ecis2024/track03_ai/track03_ai/12 (accessed on 20 January 2026).
24. Abbasi, A.F.; Chandio, S.M. Mitigating Bias in AI Systems: A Comprehensive Review of Sources and Strategies. *J. Inf. Commun. Technol. Robot. Appl.* **2024**, *15*, 45–55. [CrossRef]
25. Kusche, I. Possible harms of artificial intelligence and the EU AI act: Fundamental rights and risk. *J. Risk Res.* **2024**, 1–14. [CrossRef]
26. Veale, M.; Zuiderveen Borgesius, F. Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Comput. Law Rev. Int.* **2021**, *22*, 97–112. [CrossRef]
27. Calderon, V. Unintentional Algorithmic Discrimination: How Artificial Intelligence Undermines Disparate Impact Jurisprudence. *Duke L. Tech. Rev.* **2024**, *24*, 28.
28. Benerofe, S. AI Governance and the Verification Gap: A Framework for Law and Policy Under Computational Intractability. *SSRN* 2025, 1–73. [CrossRef]
29. Fessenko, D.S.; Jasperse, A. Ethics at the heart of AI regulation. *AI Ethics* **2025**, *5*, 3387–3398. [CrossRef]
30. Ebers, M. Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act. *Eur. J. Risk Regul.* **2024**, *16*, 684–703. [CrossRef]
31. Edwards, L. Expert Opinion: Regulating AI in Europe: Four Problems and Four Solutions. 2022. Available online: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf (accessed on 17 January 2026).
32. Kiseleva, A.; Kotzinos, D.; De Hert, P. Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations. *Front. Artif. Intell.* **2022**, *5*, 879603. [CrossRef]
33. Smuha, N.A.; Ahmed-Rengers, E.; Harkens, A.; Li, W.; MacLaren, J.; Piselli, R.; Yeung, K. How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. *SSRN Electron. J.* 2021, 5–64. [CrossRef]
34. Makauskaite-Samuole, G. Transparency in the Labyrinths of the EU AI Act: Smart or Disbalanced? *Access Justice East. Eur.* **2025**, *8*, 1–31. [CrossRef]
35. Wachter, S. Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. *Yale J. L. Tech.* **2024**, *26*, 671. [CrossRef]
36. Grozdanovski, L.; De Cooman, J. Forget the Facts, Aim for the Rights! On the Obsolescence of Empirical Knowledge in Defining the Risk/Rights-Based Approach to AI Regulation in the European Union. *Rutgers Comput. Technol. Law J.* **2023**, *49*, 207–330. Available online: https://hdl.handle.net/2268/307355 (accessed on 17 January 2026).
37. Leiser, M. Psychological Patterns and Article 5 of the AI Act. *AIRe* **2024**, *1*, 5–23. [CrossRef]
38. Novelli, C.; Casolari, F.; Hacker, P.; Spedicato, G.; Floridi, L. Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity. *Comput. Law Secur. Rev.* **2024**, *55*, 106066. [CrossRef]
39. Metikos, L. The AI Act: Weak, Weaker, Weakest. *SSRN Electron. J.* **2024**. [CrossRef]

40. EDRi. #ProtectNotSurveil: The EU AI Act Fails Migrants and People on the Move. European Digital Rights (EDRi). Available online: https://edri.org/our-work/protect-not-surveil-eu-ai-act-fails-migrants-people-on-the-move/ (accessed on 17 January 2026).
41. Clapp, S. Defence and Artificial Intelligence. 2025. Available online: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)769580 (accessed on 17 January 2026).
42. Peltner, J.; Becker, C.; Wicherski, J.; Wortberg, S.; Aborageh, M.; Costa, I.; Ehrenstein, V.; Fernandes, J.; Heß, S.; Horváth-Puhó, E.; et al. The EU project Real4Reg: Unlocking real-world data with AI. *Health Res. Policy Syst.* **2025**, *23*, 27. [CrossRef]
43. Cihon, P. Standards for AI governance: International standards to enable global coordination in AI research development. *Future Humanit. Inst.* **2019**, *40*, 340–342.
44. Pennesi, F. Equivalence in the area of financial services: An effective instrument to protect EU financial stability in global capital markets? *Common Mark. Law Rev.* **2021**, *58*. Available online: https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/58.1/COLA2021003 (accessed on 17 January 2026).
45. Balcioglu, Y.S.; Celik, A.; Altindag, E. A Turning Point in Ai: Europe's Human-centric Approach to Technology Regulation. *J. Responsible Technol.* **2025**, 100128. [CrossRef]
46. Novelli, C.; Hacker, P.; Morley, J.; Trondal, J.; Floridi, L. A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities. *Eur. J. Risk Regul.* **2024**, *16*, 566–590. [CrossRef]
47. Hong, S.H. AI and bias. In *Handbook on Public Policy and Artificial Intelligence*; Edward Elgar Publishing: Worcestershire, UK, 2024; pp. 109–122.
48. Kordzadeh, N.; Ghasemaghaei, M. Algorithmic bias: Review, synthesis, and future research directions. *Eur. J. Inf. Syst.* **2022**, *31*, 388–409. [CrossRef]
49. Tejani, A.S.; Ng, Y.S.; Xi, Y.; Rayan, J.C. Understanding and mitigating bias in imaging artificial intelligence. *Radiographics* **2024**, *44*, e230067. [CrossRef]
50. Mittermaier, M.; Raza, M.M.; Kvedar, J.C. Bias in AI-based models for medical applications: Challenges and mitigation strategies. *NPJ Digit. Med.* **2023**, *6*, 113. [CrossRef]
51. Wachter, S.; Mittelstadt, B. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI (October 5, 2018). *Columbia Bus. Law Rev.* **2019**. Available online: https://ssrn.com/abstract=3248829 (accessed on 17 January 2026).
52. Mittelstadt, B.; Wachter, S.; Russell, C. The Unfairness of Fair Machine Learning: Levelling down and strict egalitarianism by default. *arXiv* **2023**, arXiv:2302.02404. [CrossRef]
53. Balasundaram, S.; Venkatagiri, S.; Sathiyaseelan, A. Using AI to enhance candidate experience in high volume hiring: A conceptual review and case study. In Proceedings of the Replenish, Restructure Reinvent: Technology Fueled Transformation for Sustainable Future, New Delhi, India, 21–22 January 2022; pp. 21–22.
54. Fabris, A.; Baranowska, N.; Dennis, M.J.; Graus, D.; Hacker, P.; Saldivar, J.; Borgesius, F.Z.; Biega, A.J. Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Trans. Intell. Syst. Technol.* **2025**, 16, 1–54. [CrossRef]
55. Yusifli, Z. The Bouncing Ball Effect of the EU Artificial Intelligence Act on Employment Relations. *J. AI Law Regul.* **2024**, *1*, 228–232. [CrossRef]
56. De Stefano, V.; Taes, S. Algorithmic Management and Collective Bargaining. Foresight Brief No. 10. European Trade Union Institute. 2021. Available online: https://www.etui.org/publications/algorithmic-management-and-collective-bargaining (accessed on 17 January 2026).
57. Kaminski, M.E.; Malgieri, G. The Right to Explanation in the AI Act. 2025. Available online: https://ssrn.com/abstract=5194301 (accessed on 17 January 2026).
58. Minotakis, A. Regulating AI in the workplace: A critique of the EU AI act and the platform work directive through a worker-centred lens. *Platforms Soc.* **2025**, *2*. [CrossRef]
59. Addy, W.A.; Ajayi-Nifise, A.; Bello, B.G.; Tula, S.T. AI in credit scoring: A comprehensive review of models and predictive analytics. *Glob. J. Eng. Technol. Adv.* **2024**, *18*, 118–129. [CrossRef]
60. de Castro Vieira, J.R.; Barboza, F.; Cajueiro, D.; Kimura, H. Towards Fair AI: Mitigating Bias in Credit Decisions-A Systematic Literature Review. *J. Risk Financ. Manag.* **2025**, *18*, 228. [CrossRef]
61. European Commission. *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act)*; European Commission: Geneva, Switzerland, 2025; pp. 11, 50, 53, 57, 59, 73–106. Available online: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act (accessed on 17 January 2026).
62. Judgment of the Court of Justice of 7 December 2023, Schufa Holding (Scoring), c-634/21, eu:c:2023:957. paras. 14, 47. Available online: https://www.tlt.com/insights-and-events/insight/schufa-case-cjeu-rules-on-scope-of-article-22 (accessed on 17 January 2026).
63. Nathan, G. Scoring the European citizen in the AI era. *Comput. Law & Secur. Rev.* **2025**, *57*, 106130. [CrossRef]

64. Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on Credit Agreements for Consumers and Repealing Council Directive 87/102/EEC. OJ L 133, 22.5.2008, pp. 66–92. Available online: https://eur-lex.europa.eu/eli/dir/2008/48/oj/eng (accessed on 17 January 2026).
65. European Banking Authority. Final Report—Guidelines on Loan Origination and Monitoring. EBA/GL/2020/06. 2020. Available online: 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 (accessed on 17 January 2026).
66. Al-Nafjan, A.; Aljuhani, A.; Alshebel, A.; Alharbi, A.; Alshehri, A. Artificial intelligence in predictive healthcare: A systematic review. *J. Clin. Med.* **2025**, *14*, 6752. [CrossRef]
67. Kaul, T.; Damen, J.A.; Wynants, L.; Van Calster, B.; van Smeden, M.; Hooft, L.; Moons, K.G. Assessing the quality of prediction models in health care using the Prediction model Risk Of Bias ASsessment Tool (PROBAST): An evaluation of its use and practical application. *J. Clin. Epidemiol.* **2025**, *181*, 111732. [CrossRef]
68. Ebers, M. AI Robotics in Healthcare Between the EU Medical Device Regulation and the Artificial Intelligence Act. *Oslo Law Rev.* **2024**, *11*, 1–12. [CrossRef]
69. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (OJ L 117, 5.5.2017, p. 1)
70. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (OJ L 117, 5.5.2017, p. 176).
71. Leeuwen, D.L.; Gelderblom, E. The AI Act: Responsibilities and obligations for healthcare professionals and organizations. *Diagn. Interv. Radiol.* **2025**, *ahead of print*. [CrossRef]
72. Djeffal, C.; Mehl, P.; Müller, V. The EU AI Act's Impacts on Digital Health. *Curr. Dir. Biomed. Eng.* **2024**, *10*, 191–195. [CrossRef]
73. Aniela, K.; Ward, T.; Loughran, R.; McCaffery, F. Challenges Associated with the Adoption of Artificial Intelligence in Medical Device Software. In *Artificial Intelligence and Cognitive Science (AICS 2022)*; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2023; pp. 163–174. [CrossRef]
74. van Kolfschooten, H.; van Oirschot, J. The EU Artificial Intelligence Act (2024): Implications for healthcare. *Health Policy* **2024**, *149*, 105152. [CrossRef]
75. van Oirschot, J.; Ooms, G. Interpreting the EU Artificial Intelligence Act for the Health Sector. Health Action International. 2022. Available online: https://haiweb.org/wp-content/uploads/2022/02/Interpreting-the-EU-Artificial-Intelligence-Act-for-the-Health-Sector.pdf (accessed on 17 January 2026).
76. Mirindi, D. A review of the advances in artificial intelligence in transportation system development. *J. Civil Constr. Environ. Eng.* **2024**, *9*, 72–83. [CrossRef]
77. Taha, A.M.; Alkayyali, Z.K.; Zarandah, Q.M.; Abu-Naser, S.S. The Evolution of AI in Autonomous Systems: Innovations, Challenges, and Future Prospects. Int. J. Acad. Eng. Res. **2024**, *8*, 1–7.
78. Nascimento, A.M.; Vismari, L.F.; Molina, C.B.S.T.; Cugnasca, P.S.; Camargo, J.B.; de Almeida, J.R.; Inam, R.; Fersman, E.; Marquezini, M.V.; Hata, A.Y. A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4928–4946. [CrossRef]
79. Maskur, M.A.; Masyhar, A.; Damayanti, R.; Ramada, D.P.; Sanyal, S. Reimagining Criminal Liability in the Age of Artificial Intelligence: Toward a Comparative and Reform-Oriented Legal Framework. *J. Law Leg. Reform* **2025**, *6*, 1805–1838. [CrossRef]
80. Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, amending Regulation (EU) 2018/858 of the European Parliament and of the Council and repealing Regulations (EC) No 78/2009, (EC) No 79/2009 and (EC) No 661/2009 of the European Parliament and of the Council and Commission Regulations (EC) No 631/2009, (EU) No 406/2010, (EU) No 672/2010, (EU) No 1003/2010, (EU) No 1005/2010, (EU) No 1008/2010, (EU) No 1009/2010, (EU) No 19/2011, (EU) No 109/2011, (EU) No 458/2011, (EU) No 65/2012, (EU) No 130/2012, (EU) No 347/2012, (EU) No 351/2012, (EU) No 1230/2012 and (EU) 2015/166 (OJ L 325, 16.12.2019, p. 1).
81. Llorca, D.F.; Hamon, R.; Junklewitz, H.; Grosse, K.; Kunze, L.; Seiniger, P.; Swaim, R.; Reed, N.; Alahi, A.; Gómez, E.; et al. Testing autonomous vehicles and AI: Perspectives and challenges from cybersecurity, transparency, robustness and fairness. *arXiv* **2024**. [CrossRef]
82. Llorca, F.; Gómez, D. *Trustworthy Autonomous Vehicles: Assessment Criteria for Trustworthy AI in the Autonomous Driving Domain*; EU Science Hub: Luxembourg, 2021. [CrossRef]
83. Lölfing, N. Impact Of The EU's AI Act Proposal On Automated And Autonomous Vehicles–Conventus Law. Conventus Law. 2023. Available online: https://conventuslaw.com/report/impact-of-the-eus-ai-act-proposal-on-automated-and-autonomous-vehicles/ (accessed on 18 January 2026).

84. Güçlütürk, O.G.; Vural, B. Driving Innovation: Navigating the EU AI Act's Impact on Autonomous Vehicles. 2024. Available online: https://www.holisticai.com/blog/driving-innovation-navigating-eu-ai-acts-impact-on-autonomous-vehicles (accessed on 18 January 2026).
85. Vellinga, N.E. Trustworthy AI on the Road. In *Transport Transitions: Advancing Sustainable and Inclusive Mobility*; McNally, C., Carroll, P., Martinez-Pastor, B., Ghosh, B., Efthymiou, M., Valantasis-Kanellos, N., Eds.; TRAconference 2024; Lecture Notes in Mobility; Springer: Cham, Switzerland, 2025. [CrossRef]
86. Abed, A.A.; Rahma, A.M.S.; Dawood, O.A. Advancements in artificial intelligence for biometric system a systematic review. In *IET Conference Proceedings CP906*; The Institution of Engineering and Technology: Stevenage, UK, 2024; Volume 2024.
87. Michael, K.; Abbas, R.; Jayashree, P.; Bandara, R.J.; Aloudat, A. Biometrics and AI bias. *IEEE Trans. Technol. Soc.* **2022**, *3*, 2–8. [CrossRef]
88. Kindt, E.J. EU biometric data regulation: Part 2: The AI Act. IEEE Biometrics Council Newsletter. 2025. Available online: https://hdl.handle.net/1887/4273636 (accessed on 18 January 2026).
89. Kyriakides, N.; Plevri, A.; Zentani, Y. AI and access to justice: An expansion of Adrian Zuckermans findings. In *Frontiers in Civil Justice*; Edward Elgar: Worcestershire, UK, 2022; pp. 121–141. [CrossRef]
90. Mensah, G.B. Artificial intelligence and ethics: A comprehensive review of bias mitigation, transparency, and accountability in AI Systems. *Preprint* **2023**, *10*.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.