Cambridge Forum on AI: Law and Governance (2025), **1**, e24, 1–22
doi:10.1017/cfl.2025.9

Cambridge Forum

RESEARCH ARTICLE

Toward empowering AI governance with redress mechanisms

Yulu Pi[1] and Maddie Proctor[2]

[1]Centre for Interdisciplinary Methodologies, University of Warwick, Coventry, GB, UK and [2]Social Studies, Harvard University, Cambridge, USA
Corresponding author: Yulu Pi; Email: yulu.pi@warwick.ac.uk

(Received 30 October 2024; revised 5 March 2025; accepted 16 March 2025)

Abstract
Amid the rapidly evolving landscape of artificial intelligence (AI) regulation, a significant concern has emerged regarding the predominant focus on preemptive measures aimed at preventing or mitigating potential AI-related harms. While these preemptive measures are undeniably important, they must be complemented by effective redress mechanisms to address and remedy materialized harms. This paper highlights the crucial role of redress in empowering individuals to challenge and rectify the adverse effects of AI systems, emphasizing that access to redress is, in fact, access to justice. We critically evaluate whether current AI governance sufficiently address the need for remedies for AI-related harms, arguing that they fall short in protecting individuals' rights. To address this gap, we outline four key steps in the redress process: (1) initiating the redress process, (2) determining appropriate avenues for redress, (3) collecting evidence to support claims and (4) receiving and responding to decisions. Each step is explored in detail, presenting distinct challenges and requirements, illustrated with real-world examples. Our findings underscore the urgent need to integrate robust redress pathways into AI governance frameworks to safeguard individual rights as AI technologies become increasingly embedded in society.

Keywords: AI governance; redress; individual affected; ex-post measures

1. Introduction

In the face of tremendous socio-technical changes and well-documented instances of harms and risks stemming from the development and deployment of artificial intelligence (AI) (Obermeyer et al., 2019; Greene, 2023), there is a worldwide consensus regarding the necessity for regulatory intervention to address AI-related harms. Major global players, including the European Union (Union, 2021), the United States (House, 2023) and the United Kingdom (UK Department for Science and Technology, 2023), have issued their approaches for AI regulation. Although there is no agreement yet on how to regulate AI, these policies and legislative proposals have uniformly sought to emphasize regulation focusing on protecting human civil rights.[1]

[1]"The aim of the new rules is to foster trustworthy AI in Europe and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models" (European Commission, 2021). "… ensure that AI benefits the whole world, rather than exacerbating inequities, threatening human rights, and causing other harms" (The White House, 2023). "We have made an initial assessment of AI-specific risks and their potential to cause harm, with reference in our analysis to the values that they threaten if left unaddressed. These values include safety, security, fairness, privacy and agency, human rights, societal well-being and prosperity" (UK Government, 2023).

While representing a step in the right direction, critics argue that existing regulatory initiatives, such as the EU AI Act's risk-based approach and the UK's pro-innovation and cross-sectoral strategies, are insufficient in safeguarding the rights of individuals directly impacted by AI technologies (Dunlop, 2023; Rights, 2023). The calls for ensuring the rights and redress for people affected by AI systems has gained significant traction, led by human rights organizations, consumer groups, research institutions and researchers (Rights, 2023; Stockhem, 2023). This paper contributes to the ongoing discourse by advocating for robust redress mechanisms and providing a practical analysis of obstacles that hinder effective redress for those affected by AI systems. By identifying and analyzing the specific challenges encountered at each step of the redress process, this research aims to provide insights into how redress mechanism should be designed in order to provide meaningful recourse for those adversely affected by AI technologies. We stress that AI governance must protect the rights and interests of all individuals impacted by AI systems, proactively anticipating how to safeguard against harms caused by AI and providing redress when harms have materialized (Goud et al., 2023). Through a practical examination of the challenges faced in obtaining redress, this paper highlights the need for a regulatory approach that encompasses the entire AI life cycle. We posit that effective AI governance must go beyond harm prevention; it must establish comprehensive safeguards to address and rectify both individual and collective harms caused by AI. This can only be accomplished through two critical actions: (1) establishing robust mechanisms for redress that provide individuals and communities with formal avenues to seek compensation or corrective measures and (2) ensuring that these redress mechanisms are accessible and attainable for all those adversely affected by AI systems. By addressing these essential criteria, we can foster a more equitable and responsible AI ecosystem that prioritizes justice and accountability for all stakeholders.

Redress, along with related concepts such as remedy, recourse and contestability, serves as a vital mechanism for empowering individuals impacted by AI systems to protect their rights, particularly in cases where the principles of equality, inclusiveness and fairness are violated or undermined (Alfrink et al., 2022; Fanni et al., 2023). The discourse surrounding redress in addressing AI harms is frequently linked with the concept and practice of contesting AI-driven decisions (Lyons et al., 2021; Fanni et al., 2023). For instance, the concept of contestability and redress are coupled together to form the fifth principle of UK pro-innovation approach to AI regulation, which states "where appropriate, users, impacted third parties and actors in the AI life cycle should be able to contest an AI decision or outcome that is harmful or creates material risk of harms." Distinguishing the nuanced differences between these terms is beyond the scope of this paper; therefore, we use them interchangeably under the umbrella term "redress." We recognize that their common objective is to enable affected people to contest decisions, alter outcomes and seek compensation or reparation when they believe they have been treated unfairly, inaccurately or inappropriately by AI systems.

Redress refers to a range of measures aimed at addressing detriment or negative impacts experienced by individuals or communities as a result of wrongdoing (Scottish Public Services Ombudsman, 2017). The goal of redress is to provide "a remedy or set right an undesirable or unfair situation" (Fanni et al., 2023). Affected individuals can seek these measures through various avenues, including judicial mechanisms (domestic or regional courts, international human rights bodies), state-based non-judicial mechanisms (regulators, ombudsman, public complaints handling bodies) and company-based internal complaints mechanisms (United Nations Human Rights Office of the High Commissioner, 2023b). Although people might associate redress with monetary compensation, the outcome of a redress process can take different forms, including restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition (United Nations Human Rights Office of the High Commissioner, 2023b). In the context of AI unfairness, Robert et al. distinguish two types of redress outcomes: restorative and retributive (Robert et al., 2020). Restorative redress focuses on "making the offended party or the victim whole again," while retributive redress involves penalizing the offender, often through legal action. Access to redress mechanisms facilitates comprehensive investigations into human rights violations and harms, enabling the appropriate rectification of

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

harm, compensation for victims and accountability for those responsible (United Nations Human Rights Office of the High Commissioner, 2023a). Importantly, redress extends beyond the healing of individual victims and encompasses the healing and reconciliation of societies toward truth and justice (Redress, 2024a).

This paper first provides an overview of the current AI governance landscape, highlighting the disproportionate emphasis on ex-ante approaches that seek to prevent or mitigate the potential harms associated with AI. It emphasizes the critical need for mandated access to redress mechanisms within the evolving AI regulatory frameworks to address harms that ex-ante measures fail to address. It then outlines the steps necessary for seeking redress, including initiating the redress process, determining appropriate avenues for redress, collecting evidence to support claims and receiving responding to decisions. Each of these steps presents unique requirements and challenges, which are illustrated through many real-world examples. In its concluding remarks, the paper provides recommendations tailored to different actors and underscores the importance of interdisciplinary research collaboration to effectively design and enforce equitable access to redress. By positioning redress as a crucial yet missing aspect of AI governance and analyzing the barriers to obtaining it, this work seeks to prioritize the voices and experiences of those adversely affected by AI-related harms in ongoing AI governance efforts.

2. Redress mechanisms: Significance and current landscape

The right to redress is a fundamental human right, essential for addressing violation of various rights and freedom (United Nations, 1948; European Convention on Human Rights, 1950; United Nations Human Rights Office of the High Commissioner, 2023b). The significance of redress mechanisms is enshrined in international human rights instruments, such as Article 8 of the Universal Declaration of Human Rights, Article 2 of the International Covenant on Civil and Political Rights and Article 47 of the EU Charter of Fundamental Rights (United Nations Human Rights Office of the High Commissioner, 2023b). Yet, the exact applicability of these human rights instruments to AI remains unclear. Moreover, current AI governance frameworks, which primarily focus on ex-ante measures to prevent and mitigate harms, fail to adequately protect this right, leaving those affected vulnerable to harm without effective redress mechanisms. In this section, we aim to clarify the persistent misfocus in the current landscape despite ongoing regulatory efforts to address harms associated with AI.

For AI governance to safeguard individuals and society from the potential harms caused by AI, two primary approaches can be employed: ex-ante and ex-post. The distinction between ex-ante and ex-post mechanisms in AI regulation is a matter of debate. Some define the difference based on the timing of deployment: ex-ante mechanisms are forward-looking tools that take effect before an AI system is deployed and begins to impact users, while ex-post mechanisms are applied after the system has been deployed and is operational (Ada Lovelace Institute, 2021). Others argue that the distinction is more accurately understood in terms of when AI-related harms occur. The ex-ante approach focuses on preventing and mitigating harms before they occur, while the ex-post approach addresses the consequences of harms after they have taken place (Wendehorst, 2020). In this paper, we adopt the latter definition. Many AI-related harms occur not only during deployment but also during the development process. By focusing on when AI-related harms occur, the second definition allows us to reveal and address issues that may arise before deployment, such as intellectual property infringement, poor labor conditions or adverse environmental impacts. For instance, the AI Act, which adopts a product safety perspective (Almada & Petit, 2023), imposes a set of requirements ranging from data governance to human oversight for high risk AI systems to comply with before they are allowed to enter the EU market. This is an example of an ex-ante approach. However, even after meeting these requirements, AI systems may still pose the risk of causing harm. There are recognized harms and risks associated with the use of AI in various domains. Such harms include unwarranted surveillance (Greene, 2023), discriminatory practices, unjust and incorrect decision-making in the

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

areas of housing (Johnson, 2023), criminal justice (Julia et al., 2016), health care (Obermeyer et al., 2019) and many other areas. What's worse, such harms often disproportionately impact those at the margins of society, such as immigrants and those in the criminal justice system (Jones, 2023). Unfortunately, individuals who fall victim to wrongdoing, harm or unfair treatment through AI may find themselves powerless without effective avenues for reporting and addressing these harms. It is acknowledged that "many sets of AI governance principles in fact have no provision for remedy (Jones, 2023)."

The primary focus of current AI governance frameworks on ex-ante measures, while neglecting ex-post measure, particularly redress mechanisms, disproportionately impacts historically marginalized communities, exacerbating existing inequalities and injustices (Brookings, 2023). The process of reporting harms and seeking redress often requires a significant investment of time, money, and technical and legal expertise, which can be prohibitive, if not impossible, for many marginalized individuals, creating what O'Neil calls "feedback loops" (O'Neil, 2016).This feedback loop perpetuates the cycle of inequality and injustice, amplifying an already established pattern of disadvantage and making it even more difficult for these communities to challenge and change the status quo. The emphasis on the importance of redress mechanism as an ex-post measure in addressing AI-related harms is not intended to diminish the legitimacy and significance of ex-ante measures. Rather, it serves to underscore the necessity of a holistic approach to safeguarding society from the potential risks and consequences associated with AI. Ex-ante measures, such as risk assessment (National Institute of Standards and Technology, 2023), ethical guidelines (Organisation for Economic Co-operation and Development, 2019; United Nations Educational and Organization, 2021) and technical standards for design and development (International Organization for Standardization, 2021), play a crucial role in understanding, preventing and mitigating AI-related harms. These measures help to ensure that AI systems are designed, developed and deployed in a responsible manner, aiming to minimize the likelihood of harmful outcomes. However, it is also essential to recognize that even the most stringent ex-ante measures may not entirely eliminate the risk of AI-related harms. Unforeseen circumstances, emergent properties (Wei et al., 2022), malicious actors (Brundage et al., 2018) or negligence to deploy AI properly may still lead to adverse outcomes. In such cases, redress mechanisms become indispensable for affected individuals to address the consequences of these harms and holding responsible parties accountable.

Reporting and documenting AI incidents (Committee, 2023), including misuse of AI, harmful post-deployment events and the real-world consequences that follow, has gained traction as an essential ex-post measure in AI governance (Ada Lovelace Institute, 2021; Rishi Bommasani et al., 2024). These reports of AI incidents serve a critical function in ensuring that emerging risks are communicated effectively to relevant authorities and stakeholders, thereby supporting the development of sound AI policy. A survey of professionals involved in algorithmic audits found that systematic harm incident reporting ranked as the third-highest priority for regulatory intervention (Ada Lovelace Institute, 2021). By capturing instances where AI systems have been applied irresponsibly or have led to harmful outcomes, AI incident reporting enables proactive intervention and better-informed decision-making. This systematic approach allows regulatory bodies and policymakers to identify risk vectors, discern patterns of misuse, anticipate potential threats and act swiftly to mitigate risks before they escalate into broader societal harms (Kolt et al., 2024).

However, mere reporting is not enough. While it is crucial to detect and document adverse events, these actions alone do not guarantee that affected individuals or communities will receive the protection or remedies they deserve. The process of reporting must be closely linked to mechanisms for prevention and redress. Without these mechanisms, the collection of data on AI misuse remains a passive exercise – useful for analysis but ineffective in addressing the harm already caused or preventing future incidents. For reporting to have real impact, it must be part of a broader system that ensures accountability and redress. This includes creating pathways for those harmed by AI to contest decisions, correct inaccuracies and seek compensation or reparations. Following the reporting of an

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

adverse event, the processes to address the harm must be transparent and effectively communicated to all relevant parties. Such transparency not only fosters accountability but also builds trust in the reporting mechanisms themselves. By linking reporting to concrete steps for redress, these systems can hold AI developers and operators accountable for the harm their technologies may cause, ultimately contributing to a safer and more responsible AI ecosystem.

Recognizing this important gap, recent legislative updates – including the final version of the EU AI Act and the UK AI Regulation White Paper – have significantly enhanced the focus on ex-post protection and redress mechanisms in response to AI-related harms. Initially, the EU AI Act, introduced in April 2021, faced criticism for not adequately addressing human rights concerns due to the lack of a robust complaint and redress mechanism (Dunlop, 2023; Rights, 2023; Stockhem, 2023). However, subsequent revisions under the Council and Parliament proposals have made notable improvements (Engler, 2023b). The revised proposal now includes individual rights. Specifically, Articles 85, 86 and 99(10) have been added to guarantee these rights. Article 85 allows individuals or groups to file complaints with market surveillance authorities if their rights under the regulation are infringed by an AI system. Article 86 ensures the right to an explanation of the output from high-risk AI systems that impact legal rights, health, safety, socio-economic status or other fundamental rights. Article 99(10) provides for effective judicial remedies and due process against market surveillance authorities' actions. These rights were not present in the initial draft, marking a significant step toward effective individual redress. Likewise, the AI Regulation White Paper, "proportionate and pro-innovation regulatory framework" released by the UK government on March 29, 2023, emphasizes contestability and redress as crucial principles (UK Department for Science and Technology, 2023).

Despite these positive developments, establishing an effective redress mechanism requires a clear understanding of its essential components and the necessary steps involved. Many consider explainability and transparency as crucial prerequisites, since the opaque nature of AI can impede individuals in challenging decisions. However, how to achieve explainability and transparency to enable redress remains unclear (Kluttz et al., 2020; Ploug & Holm, 2020). The UK government's AI Regulation White Paper highlights contestability and redress as pivotal principles. Nevertheless, the UK's initial Guidance for Regulators published in February 2024 indicates that existing technical standards and examples of guidance or best practices for contestability and redress are notably limited compared to other principles (UK's Department for Science and Technology, 2024). Governments are grappling with questions such as: What existing routes do end users or anyone impacted by AI system, have to contest outcomes? Are these routes to contestability appropriate in the context of AI (UK's Department for Science and Technology, 2024)?

Given the critical importance of redress mechanisms in protecting affected individuals and the existing ambiguity around their implementation, the following sections will outline the essential steps for seeking redress. Each step will be examined in terms of the challenges and obstacles that individuals may face throughout the process.

3. Steps toward seeking redress for AI-induced harms

Seeking redress for harms caused or induced by AI involves navigating a complex and multifaceted process, where each step presents its own set of requirements and challenges. These hurdles can significantly hinder an individual's ability to address AI-caused harms. The process typically begins with initiating the redress process, followed by determining the appropriate avenues for seeking remedy. Subsequently, individuals must collect and present evidence to support their claims, and finally, they will receive and respond to decisions made regarding their case. These steps, while outlined as distinct steps for clarity, do not always occur in a strict sequence in real life. For example, an individual might start the redress process and simultaneously gather evidence to support their claim, or new information might prompt a reevaluation of earlier decisions about the appropriate avenues for redress. This iterative and nonlinear progression reflects the practical realities of addressing AI-induced harms.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

3.1. Step 1: Initiate the redress process

The first step in seeking redress is initiating the process, which begins with identifying the specific AI systems involved and understanding the harms they've caused. However, this step is often challenging due to the lack of transparency surrounding AI use and the delayed recognition of its negative impacts. Transparency in the use of AI is essential for helping individuals understand their interactions with AI systems, whether directly or indirectly, and for addressing any potential harms or unfair treatments that may arise. A distinct challenge is that individuals are often unaware that they have been subjected to AI-driven decision-making processes or have encountered AI-generated content (Goud et al., 2023). AI systems are frequently deployed without public notice, making it difficult for affected individuals to identify these systems and understand their consequences. This unawareness is particularly problematic in cases of faulty decisions or institutional transgressions (Selander et al., 2023). It significantly hinders individuals' ability to recognize wrongdoing or unfair treatment, thereby hindering the initiation of the redress process. For example, the City of Pittsburgh utilized a predictive policing algorithm for over a year before the public was informed of its existence. The algorithm's use began in February 2017, but it was not until October 2018, through a paper published by CMU researchers, that the public became aware of its deployment (Johnson et al., 2024). This lack of transparency delayed the recognition of potential harms and the initiation of any redress process. The AI Act, now enacted, attempts to address such transparency challenges by imposing various obligations on providers and deployers of AI systems or general-purpose AI models. These obligations encompass maintaining technical documentation, conducting risk assessments and ensuring traceability. Specifically, Article 13 requires that high-risk AI systems be designed and developed with sufficient transparency, allowing deployers to interpret the system's outputs and use them correctly. These provisions seek to ensure that users are aware when they interact with an AI system, inform them of the system's capabilities and limitations and notify affected individuals of their rights. Enforcing these transparency obligations is crucial for the effectiveness of redress mechanisms. It requires ensuring that users and affected individuals are fully aware when they are interacting with an AI system or when they are subject to AI-driven decisions. This awareness is crucial because, without it, people may not recognize when an AI system has caused harm, making it difficult for them to initiate the process of seeking redress.

Another significant challenge to initiate the redress process is the delayed recognition of AI-induced harms, particularly those that are cumulative and collective in nature (Smuha, 2021). Such harms may not become apparent until they have accumulated to a substantial level. AI-related harms like discrimination and unfairness often manifest more clearly when viewed collectively, rather than in isolation. At an individual level, a single instance of bias or error in an AI system might seem minor. However, the widespread use of such systems can infringe upon the rights of many individuals, particularly affecting vulnerable and marginalized groups. The case of the Equal Employment Opportunity Commission (EEOC) versus iTutorGroup exemplifies this delayed recognition of AI-caused harms. iTutorGroup's algorithmic hiring practices systematically excluded applicants aged 65 and older (U.S. Equal Employment Opportunity Commission, 2024). Initially, the discriminatory nature of this practice appeared trivial, as individual rejections seemed isolated. Over time, it became clear that the algorithm had consistently rejected over 200 qualified applicants based solely on age, revealing a significant pattern of bias. This delayed recognition of harm complicates the initiation and progression of the redress process, as the full extent of the damage may only become apparent after a considerable delay, leaving many more individuals exposed to such harms.

3.2. Step 2: Determine appropriate avenues for redress

Once the harms have been identified, the next step is to determine the most suitable avenues for seeking redress. These can include company-based internal complaints mechanisms, state-based non-judicial mechanisms such as ombudsman services and judicial mechanisms like courts or

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

tribunals (Business and Human Rights Resource Centre, 2016; Stefan Zagelmeyer and Shemberg, 2018). This list is not exhaustive but serves as a starting point for regulators to map and assess the capacity of existing avenues to address mistakes made or harms inflicted by AI systems (Floridi et al., 2018). There are inherent challenges related to general redress process, characterized by a fragmented and inconsistent landscape, as highlighted in a recent debate on access to redress schemes in the UK (Rhodes et al., 2024). This issue is further complicated by the fact that AI's harms can span multiple sectors. Individuals need clear guidance to navigate the best avenues for their specific situations, which can be quite confusing. Below, we explore three primary mechanisms in more detail, with a specific focus on the challenges individuals may encounter when seeking redress through these avenues: internal complaint mechanisms, non-judicial dispute resolution and judicial mechanisms.

Internal complaint mechanisms

Internal complaint mechanisms are often the first point of contact for individuals seeking redress for harms caused by a product or service, including those related to AI (Digital Regulation Platform, 2020). These mechanisms provide an accessible way for people to report issues and obtain timely resolutions. An effective internal complaint system holds AI providers accountable by promptly addressing harms, fostering a culture of responsibility. Moreover, these mechanisms can yield valuable insights into the negative impacts of AI products and services, enabling companies to make meaningful improvements in service delivery. The benefits extend beyond operational enhancements; a well-functioning complaint system can also elevate an organization's reputation and build public confidence in its offerings. In this way, redress not only empowers individuals but also strengthens the company's relationship with its customers. However, it remains unclear whether company-based internal complaint mechanisms specifically addressing AI issues exist or are effective.

Several barriers can obstruct victims of AI-related harms when attempting to use these complaint mechanisms. Often, these obstacles stem from:

- **Lack of incentive:** Companies may not prioritize effective complaint channels, resulting in limited accessibility and visibility. Without a strong incentive – such as regulatory pressure, financial penalties or significant reputational damage – companies might deprioritize the development and maintenance of effective complaint mechanisms.
- **Insufficient transparency:** Individuals may not receive clear guidance on how to report issues or understand their rights to contest AI-driven decisions. If companies do not provide transparent information about how to file complaints or appeal decisions made by AI systems, affected individuals may struggle to navigate the process.

A notable incident illustrating these barriers involves the Apple Card. The Apple Card faced allegations of gender bias, with claims that it assigned different credit scores to individuals with similar or even identical financial profiles but different genders (New York State Department of Financial Services, 2021). Reports indicated that "no appeal worked" (DHH, 2019), raising doubts about the existence and effectiveness of internal compliance mechanisms. Moreover, consumer claims of AI-related harms are sometimes met with outright denial from companies. For example, when an Instagram user shared an infographic describing a troubling interaction with the National Eating Disorders Association's (NEDA) Tessa Wellness chatbot, NEDA's VP of Communications initially accused the user of lying in the comments. The backlash from other users led the VP to delete her comment and issue a retraction (Johnson et al., 2024).

Another challenge with redress through a company's internal process is that many laws and regulations related to AI issues, such as data protection and copyright infringement, are nascent and still evolving. Consequently, companies might not be fully aware of their legal obligations or the rights of affected individuals. For instance, following the implementation of the General Data Protection Regulation (GDPR) in 2018, there was significant debate about the right to explanation (Selbst and

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Powles, 2017; Wachter et al., 2017). It was not until 2021 that a court in Europe first recognized this right, when the Court required Ola, a ride-hailing platform, to explain the logic behind its automated system of penalties and deduction to its drivers (Amsterdam District Court, 2021). Before this court decision, companies could operate in a regulatory gray area where obligations were unclear, potentially hindering affected individuals from seeking redress via internal company processes. Additionally, Article 17 of the Digital Services Act (DSA) requires online platforms to establish and provide access to complaint systems, allowing customers to submit substantiated complaints about the removal or disabling of content created by them on a platform or suspension/removal of their account. Although these mechanisms have been integrated into data protection and online platform regulations, extending such protections to all AI systems remains a challenge (Ogunleye, 2022).

To overcome these barriers, a comprehensive and robust protocol for handling feedback and complaints should be implemented within AI providers and deployers. This protocol should emphasize transparency in communication, ensuring that individuals are provided with timely notifications and updates regarding their complaints. Maintaining open and honest communication is vital for building and maintaining trust between companies and those affected by AI-related harms. Additionally, sectoral regulators must take a proactive role in overseeing and enforcing effective complaint-handling protocols within companies, ensuring that these systems are not only functional but also fair and transparent.

Dispute resolution: AI ombudsman

Typically, consumers experiencing harms first bring their complaints to product or service providers through in-house complaint mechanisms. If their complaints are unresolved or they are unsatisfied with the resolution, complainants may escalate to a further stage where external dispute resolution – such as ombudspersons – get involved. An ombudsperson serves as an independent and impartial intermediary tasked with addressing disputes. Ombudsmen offer non-judicial dispute resolution schemes, valued for their efficiency, cost-effectiveness and flexibility compared to court proceedings (Vickers, 2022). Legal recourse, such as filing complaints or litigation, can be inaccessible and difficult to navigate to individuals with limited resources or time. Court and attorney's fees, months and years-long legal processes, and the delegation of resolution power to a judge or jury can discourage victims from pursuing legal redress. Resolution by ombudsman is cheaper, faster and allows parties greater agency in reaching a fair solution.

In its report "Regulating AI in the UK," Ada Lovelace Institute made recommendations establishing an "AI ombudsman" to support people impacted by AI (Davies and Birtwistle, 2023). Additionally, apart from supporting individuals in resolving their complaints and guiding them to suitable regulators when needed, AI ombudsman has a vital role in informing governments and regulators about the types of AI-related harms individuals experience and whether they effectively obtain redress to increase the visibility and awareness of AI-caused harms. An effective AI ombudsman requires expertise in AI as well as domain-specific knowledge related to disputes. Two approaches to establishing AI ombudsman with the appropriate expertise can be pursued: enhancing the capabilities of existing ombudsmen by building internal AI expertise or creating specialized AI ombudsmen who work closely with sector-specific regulators and ombudsmen. According to several EU surveys, consumers might be reluctant to pursue any litigation no matter how severe they experience financial loss – especially those who are older, less educated, live alone, are retired, widowed or do not use computers (Hodges, 2014). Particularly if the alternative is the absence of any redress, informal dispute resolution is a valuable tool for addressing complex AI harms, making the process of seeking redress more accessible to consumers who are hesitant to engage in legal action. By offering informal dispute resolution options, an AI ombudsman serves as a valuable mechanism for addressing complex AI-related harms, enabling affected individuals to seek redress without having to navigate the complexities of the legal system.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

One notable ombudsman resolution that successfully achieved redress by informing victims and preventing further privacy violations took place in Finland. In 2020, the Finnish National Bureau of Investigation used facial recognition software from Clearview AI to identify potential victims of child sexual abuse without implementing appropriate privacy safeguards, such as restrictions on data storage duration or third-party sharing. The National Police Board should have notified the Finnish Office of the Data Protection Ombudsman about the project, which they eventually did in 2021 after reporting a personal data breach. In response, the ombudsman ordered that the victims be informed of the breach and that Clearview AI erase the relevant personal data (Board, 2021).

Regulatory and consumer protection agencies for redress

Regulatory and consumer protection enforcement authorities play an important role in achieving fair outcomes for large-scale consumer harms from AI. The International Consumer Protection and Enforcement Network (ICPEN) is an international organization led by the United States' Federal Trade Commission (FTC) and composed of 70 member authorities, including the UK's Competition and Markets Authority (CMA). The European Commission is also an observer authority to ICPEN. In the US, the FTC has broad authority to seek redress for a variety of AI harms. Section 5 of the FTC Act, for example, establishes the Commission's mandate to regulate unfair and deceptive practices (Raji et al., 2022). The FTC can sue companies and refund injured parties on a pro rata basis (Federal Trade Commission, 2021), issue injunctions forcing companies to stop harmful practices or enter into long-term consent decrees allowing for future monitoring and fines (Raji et al., 2022). At a Tech Summit hosted by the FTC's Office of Technology, FTC Chair Lina Khan cited the Commission's "recent robocall enforcement sweep" as an example of how they will approach AI liability. In the robocall sweep, the FTC partnered with state attorneys general to go after upstream actors including lead generators and voice over internet protocol providers as part of "Operation Stop Spam Calls" (Federal Trade Commission, 2023). Chair Khan's use of Operation Stop Scam Calls – and the FTC targeting Walmart as an upstream payment actor in recent anti-fraud cases (Federal Trade Commission, 2021) – as models for AI liability signals the Commission's willingness to seek redress from upstream actors in particular. The FTC has broad authority to target misleading or deceptive practices, for which the FTC needs only to show that a claim about an AI product was misleading. Many US states also enforce misleading products claims, like California's Unfair Competition Law which allows individual consumers to sue for injunctive relief (Raji et al., 2022).

The scope and powers of consumer protection and market competition authorities are expanding to effectively address harms arising from digital products, including AI. For instance, the UK's Digital Markets, Competition and Consumers Bill empowers the CMA with ex-ante authority and the ability to impose tougher fines (Competition and Markets Authority, 2024). This legislative framework elevates consumer protection law to the same level as competition, significantly enhancing the CMA's capacity to safeguard consumers from commercial harm and deter businesses that do not adhere to the rules. In instances of systemic harms, consumer protection authorities can leverage their unique ability to conduct market-wide investigations, providing financial and other forms of redress. However, consumer protection powers vary internationally and represent only one facet of a fair and comprehensive redress ecosystem. In contrast, private rights of action may be more suitable for addressing individual cases and smaller-scale harms.

Moreover, jurisdictional complexities and procedural challenges remain significant, particularly in composite decision-making, where multiple authorities engage at different stages (Demková, 2023). A key issue is the limitation of judicial review, which is often restricted to final decisions while overlooking earlier preparatory stages, where data protection authorities may have an initial role under the GDPR (Demková, 2023). To address these challenges, ongoing efforts are needed to enhance access to regulatory and consumer protection agencies, ensuring effective redress mechanisms for both intermediate and final outputs of AI systems and automated decision-making.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Judicial review and litigation-based redress

Tort law offers several pathways for civil redress, including the torts of products liability and negligence. Historically, products liability has struggled to recognize software as a "product," which complicates claims for financial damages arising from digital products (DiMatteo et al., 2022). In the United States, there has yet to be a successful product liability claim involving software. In October 2024, Council of the EU approved a new "Product Liability Directive" (PLD) that extends liability for defective products to include digital products and software (European Council, 2024). This reform significantly lowers the barriers for plaintiffs across Europe to pursue product liability claims for AI-related harms. Unlike traditional fault-based liability, the PLD assumes a standard of strict liability allowing for claims without manufacturer fault or negligence (European Parliament, 2025). Additionally, the PLD introduces presumptions of defectiveness in cases where defendants fail to comply with disclosure obligations, violate legally prescribed safety requirements, or when damage results from an obvious product malfunction. While the PLD represents a significant step toward AI liability, critics argue that it remains a half-measure, failing to fully address the complexities of AI-related harms (Hacker, 2023). Many of its provisions raise difficult questions regarding interpretation, particularly in balancing trade secret protections, market competition and the effective compensation of injured parties.

Beyond product liability, negligence law remains an important legal avenue for redress. Negligence applies to developers, deployers and all participants in the AI supply chain, requiring plaintiffs to demonstrate that a provider failed to exercise "due care" in AI development and deployment (Goertzel, 2016). However, proving negligence entails significant investigative costs, which are often even higher in software-related cases. This financial burden raises fairness concerns, as the prohibitively high costs may deter victims from seeking redress through this legal avenue. Consequently, it highlights the necessity for more affordable alternatives, such as informal dispute resolution methods, outside the court system.

Additional legal avenues for redress include breach of warranty, fraud and potential strict liability claims under the doctrine of abnormally dangerous activities. Cases involving AI-related harms are likely to encounter many challenges, such as injury-in-fact, causality and redressability – similar to those faced by software and algorithms (Metcalf et al., 2023). Plaintiffs must demonstrate injury-in-fact by showing concrete and particularized harm that is actual or imminent, but algorithmic harms can be diffuse and probabilistic, complicating this process. Establishing causality is also complex, as algorithmic systems often involve multiple variables that obscure the causal chain due to system complexity or lack of developer transparency. Moreover, plaintiffs must prove that their harm can be remedied by a court order, which is challenging when the nature of the harm is not easily addressed through traditional legal remedies or when the court lacks the expertise to understand and mitigate the impact of algorithmic systems. These complexities underscore the need for further efforts to clarify unresolved legal issues regarding injury-in-fact, causality and redressability for harms related to digital products.

Collective redress

The challenge of navigating redress avenue for "cumulative" or "collective" harms arising from AI is becoming increasingly pronounced. While individual suffer from bias, discrimination or errors within AI systems which may seem insignificant, the large scale use of such systems can pose risks that infringe the rights of large numbers of people (United Nations Human Rights Office of the High Commissioner, 2023c). Compounding this problem, these cumulative impacts fall hardest on already vulnerable and marginalized populations. For instance, the generation and dissemination of deepfakes primarily target women and children, leading to increased online harassment and privacy abuses (Government of the United Kingdom, 2021). However, it is unlikely that a person harmed by AI will be able to bring cases related to collective harms to court individually due to several factors, including a lack of awareness, financial and time constraints, and the expertise required for litigation.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Given that many AI harms can only be fully understood and addressed through the lens of collective impacts, there is an urgent need to integrate clear and well-defined redress mechanisms, both for individual and collective harms, within emerging AI regulation frameworks.

In light of these challenges, representative actions provide a realistic pathway for individuals to collectively seek redress for shared harms caused by AI. This court-based mechanism allows a group of individuals, who have been similarly harmed by the same entity, to come together and pursue legal action collectively. There is a well-recognized assumption of power and information asymmetry between consumers and those providing AI, which can be mitigated by enabling individuals to seek redress as a collective. The Representative Actions Directive (RAD), adopted by the EU in November 2020, aims to empower consumers to collectively enforce their rights through redress measures. Under this directive, qualified entities can bring representative actions before national courts or administrative authorities on behalf of affected groups (European Commission, 2024b). These actions can compel product or service providers to stop illegal and harmful practices or seek compensation for affected consumers. The inclusion of the RAD in the AI Act's Annex I: List of Union Harmonisation Legislation underscores the importance of collective redress in addressing AI-related harms (European Commission, 2024a). This inclusion means that the AI Act grants consumers the right to invoke collective redress when an AI system has caused harm to a group of consumers. This not only strengthens principles of consumer protection and access to justice but also aligns with recent EU digital laws, such as the DSA, Digital Markets Act and Data Act, all of which fall under the scope of the new rules on representative actions outlined in the RAD.

In addition, collective alternative dispute resolution (ADR) offers an accessible and practical solution to fill the private enforcement gap for those impacted by AI. In collective ADR, individuals must actively opt in to participate in the dispute resolution process, which typically occurs outside the court system through methods such as mediation, arbitration or ombudsman facilitation. This approach can be particularly effective in addressing large-scale harms that are often deprioritized by consumer protection authorities, allowing for more flexible and efficient resolution of disputes. Various ADR frameworks employed by EU Member States include direct negotiations through lawyers and adjudication through arbitration tribunals. For instance, the World Intellectual Property Organization's ADR for AI disputes specifically targets issues such as copyright infringement and data privacy.

3.3. Step 3: Evidence collection to support claims for redress

Collecting evidence to substantiate claims of harm caused by AI-related harms attributable to the responsible party is a critical step in the process of seeking redress. This evidence collecting process can begin as soon as the harm has been recognized, or it may continue to be refined and directed by the identification of appropriate avenues for redress. The role of evidence in seeking redress for AI-induced harm is crucial, as it forms the foundation upon which claims are substantiated and justice is pursued. Evidence can be defined as any means used to support claims for redress, encompassing both material items and assertions of facts that help ascertain the truth of the alleged matter (REDRESS, 2024b). However, the identification, collection and presentation of evidence in cases involving AI present unique challenges that can significantly impede the pursuit of justice. We will dive into two significant challenges arise during this process: the opacity of AI systems and the "many hands" problem caused by the intricate AI value chain.

The opacity of AI

One of the primary challenges in evidence collection is the inherent technical opacity of AI systems, which can be compounded by deliberate corporate or state confidentiality and the general public's technical illiteracy (Burrell, 2016). Making the presence of AI transparent as we suggested in step 1 is not enough. Unlike traditional automated systems, such as expert systems that rely on pre-defined if-then rules, AI powered by machine learning or deep learning models – ranging from

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

early perceptrons to modern deep neural networks – is designed to "learn from experience" through iterative optimization processes (Ghahramani, 2015). This shift to data-driven learning introduces significant complexity, characterized by nonlinearity due to the activation functions that model intricate relationships within the data. Additionally, the stochastic nature of many training algorithms means that identical input data and model configurations can yield different outputs, reflecting the inherent variability in the learning process. As a result, even with the same starting conditions, an AI model may converge to different solutions, highlighting the complexities and unpredictability of these systems. Consequently, these models operate as black boxes – while we can observe their inputs and outputs, the internal mechanisms driving these processes remain hidden and incomprehensible. Moreover, the opacity of AI extends beyond technical complexities. Companies that build AI systems often cite proprietary algorithms and confidential data as justifications for withholding detailed explanations from those affected. A notable example occurred when teachers in Houston successfully sued their schools over an algorithmic system that evaluated their performance (Lutz, 2017). While high evaluations were praised, those with poor ratings faced the risk of termination. Some teachers believed they were unfairly penalized but could not verify their suspicions, as the software developer, the SAS Institute, regarded its algorithm as a trade secret and refused to disclose its workings. Ultimately, a federal judge ruled that the program had violated the teachers' civil rights when they brought their case to court. These opacity obstructs the assessment of errors or unfairness, impeding the pursuit of redress for any resulting harm.

The European Parliament addressed this challenge by incorporating a requirement into the EU AI Act, mandating the notification of individuals affected by high-risk AI systems (EU AI Act, 2024). This notification would also grant individuals the right to obtain an explanation about the role of the AI system in the decision-making procedure and the main elements of the decision taken (Engler, 2023b). Many agreed that the nature of the explanation to affected people should be concise, easily understandable and accessible (Felzmann et al., 2019). Nevertheless, the question remains: What should be included in explanations to empower individuals to grasp the workings of the AI system in order to seek remedies and redress when appropriate? While it is widely acknowledged that the answer is domain- and context-specific (Ploug & Holm, 2020), there is limited research and regulatory focus on the actions, such as contesting AI outputs or seeking redress, that explanations can enable. Ploug and Holm identified four dimensions of explainability for effective contestation in AI health diagnostics, namely the AI system's use of data, the system's potential biases, the system's performance and the division of labor between the system and health-care professionals (Ploug & Holm, 2020). There's an urgent need to comprehend how explanations can empower individuals to effectively challenge AI decisions for obtaining redress. While the provision of explanations to AI's output is recognized as prerequisite for contestation, previous studies suggest that providing too much detailed information may overwhelm users, potentially leading to confusion and reduced performance due to information overload (Oh et al., 2018; Ferguson et al., 2022).

Moving forward, to provide explanations for AI systems with the aims of identifying harms, contesting AI outputs and seeking redress, there should be a shift in the evaluation paradigm. The shift should be toward two directions: First, an emphasis on application-grounded evaluation, which evaluates the effectiveness of AI explanations in real-world contexts rather than proxy tasks in simplified experimental settings (Amarasinghe et al., 2024). Second, a focus on objective assessments of individuals' capability to identify AI errors, challenge decisions and pursue remedies instead of solely on user's perceptions, trust and understanding (Mansi & Riedl, 2023).

Many hands problem

The European Parliament's EU AI Liability Directive (AILD) underscores another significant challenge: "The large number of people potentially involved in the design, development, deployment and operation of high-risk AI systems, makes it very difficult for plaintiffs to identify the person potentially liable for damage caused and to prove the conditions for a claim for damages" (European

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Parliament, 2023). This challenge is extensively discussed in the literature as the "many hands problem" (Coeckelbergh, 2020; Cobbe et al., 2023a; Khosrowi et al., 2024). Defining accountability – who is responsible or answerable for an AI system, its behavior and its potential impacts – is crucial for those seeking redress (Raji et al., 2020). Entities involved in the development, operation, usage and monitoring of AI systems are responsible for any harm those systems may cause. Transparency serves as the foundation of accountability, where the accountable party is obligated to explain and justify their conduct (Busuioc, 2021). For individuals seeking redress after being negatively impacted, meaningful accountability requires a clear identification of those responsible and the extent of their fault for the harms caused by AI.

Establishing clear accountability, however, is fraught with challenges. First, having a clear line of accountability within a single organization for AI development and deployment face many barriers. Drage et al. (2024) conducted empirical interviews within a large multinational tech company and unveiled significant challenges in establishing clear accountability. These hurdles include high employee turnover and team transitions, a lack of incentives for routine maintenance tasks and structural barriers hindering taking ownership over AI products. Elish 2019 cautions against "Moral Crumple Zones," where responsibility for AI-induced harms is often misplaced on human users or operators who have limited control over the system's behavior. This misattribution can obscure the source of malfunctions and harm, making them not immediately apparent. Consequently, human operators may be erroneously held accountable, despite their limited ability to foresee or prevent AI errors. This dynamic enables developers and deployers to evade responsibility, leaving the human users to bear the brunt of the consequences. Moreover, the complexity deepens when considering that the design and development of AI products often transcend sectors and organizations. The difficulty arises from the intricate technological and business relationships entwined in the entire life cycle of AI systems. AI value chain involves developers, deployers, users and those directly affected, making the determination of accountability a complex endeavor. The fragmentation of control and responsibility across those actors in the AI value chain leads to a "many hands problem," where no one is responsible for outcomes which multiple people helped produce (Cobbe et al., 2023b).

To identify the source of harms, a comprehensive mapping of roles and responsibilities across stages, from data creation and curation to training, adaptation and deployment, is needed (Bommasani et al., 2021). This means that it must be evident who bears responsibility for AI's operation at all stages of the design and deployment of AI (Jones, 2023). In particular, a collaborative system is crucial for establishing clear accountability among all involved parties. In this system, AI developers (e.g. developers for a AI hiring tool) are tasked with responsibly collecting and processing data while mitigating bias. They must transparently communicate both the capabilities and limitations of their systems, providing deployers with sufficient information to conduct risk management and provide adequate explanations to affected individuals. Deployers (e.g. company using the AI hiring tool), on the other hand, are required to conduct fundamental rights impact assessments, ensuring a thorough understanding of how those systems function and establishing mechanisms for public transparency, especially for those affected by high-risk AI. Additionally, users (e.g. HR in the company) should undergo training to acquire the adequate knowledge and skills needed for the effective and responsible use of AI systems.

The AILD seeks to clarify liability for harms caused by AI, addressing challenges posed by the many hands problem and the opacity of AI systems. One of its key provisions enables national courts to mandate the disclosure of evidence related to high-risk AI systems suspected of causing harm. Companies responsible for such systems must provide various forms of documentation, including detailed logs and other relevant records (European Parliament, 2023). Another important provision introduces a rebuttable "presumption of causality" (European Parliament, 2023), which alleviates the burden of proof for claimants who face difficulties in demonstrating direct causation. If a claimant establishes that the defendant failed to comply with legal obligations – such as those imposed by the AI Act – and that this non-compliance is reasonably linked to the harm

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

suffered, the presumption operates in favor of the claimant. This mechanism is instrumental in addressing information asymmetry between AI providers and those affected by AI-related harm, thereby strengthening legal accountability and access to justice. However, the AILD was withdrawn by the European Commission in February 2025 due to a lack of consensus (European Commission, 2025). While the AILD's failure signals ongoing regulatory uncertainty, some argue that a harmonized liability regime could improve access to redress and ensure consistency across the EU.

3.4. Step 4: Receive and respond to decisions

As a result of the redress process, harmed parties will receive decisions made regarding their case. They must then evaluate the outcome and respond accordingly. If the decision is favorable and the affected parties are satisfied with the outcome, the redress process for the involving case is considered complete. However, if the individual disagrees with the decision or believes that the remedy does not adequately address harm, they should have the option to appeal the decision or seek further redress through alternative mechanisms. This could involve escalating the case within the same avenue or exploring other avenues that may offer a more favorable outcome. International law recognizes various remedies for human rights violations, which can be applied to address harms caused by AI. These remedies include restitution, compensation, rehabilitation, satisfaction and measures to prevent recurrences (United Nations Human Rights Office of the High Commissioner, 2023b).

Restitution aims to restore individuals to their pre-harm state, but achieving full restitution is often challenging or impossible, especially in cases of physical or mental harm (Gallen and Moffett, 2022). For example, the UK exam boards' decision to rescind algorithm-assigned grades and revert to teacher assessments was an important step toward correcting the harm caused. However, such remedies does not address the full extent of the damage, such as the emotional distress and lost opportunities resulting from the delay. Ehsan et al. (2022) introduced the concept of the "algorithmic imprint" to illustrate how simply removing an algorithm and attempting to revert to a pre-harm state do not necessarily undo or mitigate the lasting effects of its use.

Compensation – such as the Dutch government's €30,000 payments to families wrongly accused of benefits fraud by an algorithm (Henley, 2021) – can provide financial relief. However, monetary compensation alone fails to address the systemic and structural issues of bias and inequality embedded in and perpetuated by many AI systems. Moreover, it does not address the broader social harm that affects entire communities, which requires structural reform.

Rehabilitation seeks to restore both physical and psychological well-being, often involving mental health services for those affected by AI-related harms. However, our research shows that such support is rarely offered to individuals experiencing mental and psychological distress. Compounding this issue, the process of seeking redress for AI harms can also lead to significant frustration (Broussard, 2023). As Sara Ahmed notes in her critique of formal complaint processes, "You end up having to complain about how your complaint is handled" (Ahmed, 2021). This highlights the cyclical and bureaucratic nature of pursuing redress, where the process itself can become an added burden. Individuals often find themselves trapped in a system that complicates rather than facilitates justice, navigating multiple layers of institutional procedures that hinder their quest for redress.

Similarly, satisfaction – such as public apologies from companies like Google or Twitter for biased algorithms (Hern, 2020; Mitchell, 2024) – offers symbolic redress but often fails to address the broader structural issues that led to the harm. In cases involving systemic or collective harm, such gestures can feel inadequate, as they do not dismantle the larger systems of inequality that caused the damage in the first place.

Guarantees of non-repetition typically involve the abandonment of algorithms – an organization's decision to cease the design, development or use of a particular algorithmic system due to its potential harms (Johnson et al., 2024). While essential for preventing future damage, this approach requires fundamental changes not only to AI systems but also to the societal structures that enable algorithmic

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

harms. One-time fixes for isolated cases is often inadequate for addressing systemic harms. The same holds true for AI-related harms, where meaningful redress requires sustained efforts to dismantle entrenched social inequalities embedded in and perpetuated by algorithms. Addressing these issues goes beyond providing redress to individual victims – it involves confronting the structural inequalities that are reinforced and perpetuated by AI systems. These inequalities are tied to ongoing societal issues like racism, sexism and classism, making redress a long-term and multifaceted process. Ahmed's work on complaint processes underscores this point: seeking justice is not just about correcting individual wrongs but also about challenging the institutional structures that hinder meaningful change. As she argues, complaint processes can sometimes serve as mechanisms to delay or deflect justice, rather than resolve the underlying issues (Ahmed, 2021).

In this context, addressing AI-related harms becomes part of the broader, ongoing project of dismantling systemic injustice. The process of seeking redress is not static; it requires continuous engagement, advocacy and reform. Real progress in solving algorithmic harms involves not only responding to immediate grievances but also confronting and reshaping the societal structures that allow such harms to occur in the first place. This makes the pursuit of justice a continuous and evolving process, one that extends beyond redress for individual cases and demands long-term commitment to dismantling the societal inequalities embedded within AI systems.

4. Closing remark

This paper has stressed the significance and pressing need of building effective redress mechanisms to protect the rights of individuals and groups affected by AI. AI systems are increasingly integrated into many key sectors such as health care, finance, employment, criminal justice, education and beyond. The benefits of enhanced efficiency, accuracy and personalization come hand in hand with significant challenges related to privacy infringement, discrimination and potential violations of fundamental rights. Frequently, these negative and harmful impacts disproportionately affect the most marginalized and vulnerable individuals, groups and communities.

In the face of such harmful impacts, having redress mechanisms that empower individuals to challenge and rectify adverse AI impacts becomes essential to upholding principles of justice and equity. While global attention has been drawn to evolving regulatory frameworks that aim to minimize risks associated with AI while promoting innovation, a noticeable gap is emerging in the realm of remedy or redress. This paper highlights the challenges faced by individuals seeking redress, using real-world examples to demonstrate the various obstacles encountered at each step toward redress, primarily in the EU and US. The regulatory approaches of the EU and the US are often compared due to their distinct AI governance models – where the EU emphasizes regulatory oversight and the US tends to favor a more market-driven and decentralized approach (Engler, 2023a; Tréhu & Ricart, 2024). By examining these vastly different frameworks, it becomes clear that neither has fully addressed the need for effective remedies for individuals harmed by AI systems. The inadequacy of existing redress mechanisms exposes individuals to significant risks, whether due to data breaches, accidents involving autonomous vehicles or biases in algorithms. Redress mechanisms are vital safeguards, providing affected individuals with avenues for seeking compensation, rectification or, at the very least, human review of decisions made by AI systems. Such mechanisms not only protect individual rights but also foster trust among consumers, which, in turn, supports the long-term growth and ethical development of the AI industry.

The alarming lack of redress mechanisms in the emerging AI regulatory frameworks is now facing growing criticism that goes so far as to argue that these regulations fail to adequately protect fundamental rights (Rights, 2023). To address this gap, a combination of ex-ante and ex-post measures is essential. Ex-ante measures, already prevalent in existing AI governance initiatives such as risk management, impact assessments, design and evaluation requirements, can help prevent AI-related harms by identifying and mitigating potential risks before they materialize. These measures play a crucial

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

role in ensuring that AI systems are designed and implemented in a responsible and ethical manner. Ex-post measures, including redress mechanisms, provide effective remedies for individuals who have experienced AI-related harms. These mechanisms ensure that affected individuals receive appropriate compensation and that companies are held accountable for their actions. By implementing robust redress systems, we can address the harms caused by AI and foster trust in AI.

In this paper, we highlight four essential steps in the redress process for harms caused by AI: initiating the redress process, identifying suitable avenues for remedy, collecting evidence to support claims and responding to decisions. Seeking redress for AI-induced harms is a complex and multifaceted endeavor, with each step presenting its own challenges and requirements. These challenges can significantly impact an individual's ability to address AI-related harms. We also analyze how key AI-related regulatory measures, such as the AI Act, the AILD, product liability laws and the GDPR, aim to address these challenges. However, we acknowledge that some countries, including the US, are adopting non-binding approaches to AI governance, and not every instance of AI-related harm necessarily requires legal action to obtain redress. In this context, it is essential not to overlook the role of existing redress mechanisms, particularly those in sectors with well-established mechanisms, in adapting to the AI context. For example, many countries have financial ombudsman services or regulatory bodies that assist consumers, such as the financial ombudsmen in various EU countries and the Consumer Financial Protection Bureau in the US. These organizations offer low-cost accessible avenues for people to resolve disputes with financial institutions (UK Finance, 2021). Similarly, the health-care sector benefits from robust patient redress systems, like the UK's National Health Service (NHS) complaints procedure, which offers structured pathways for patients to complain about services, or procurement, patient choice and competition in the NHS (Government of United Kingdom, 2025). These existing mechanisms can provide valuable insights into how redress can be effectively implemented across the diverse range of AI applications.

Moreover, addressing the current shortcomings in the redress mechanisms for the impacts of AI requires a collaborative approach. It is essential for various stakeholders – such as supervisory authorities, AI developers, deployments and civil society – to work together effectively to address existing gaps and establish robust redress mechanisms.

Supervisory authorities: To effectively support affected individuals and communities, it is fundamental to provide a legally binding basis for transparency, accountability and clear routes to redress within the AI regulatory framework. This involves mandating the "notification of ability to contest," which ensures users are informed of their right to contest and seek redress (Fanni et al., 2023). Moreover, it entails establishing clear guidelines that specify the types of redress individuals can pursue, the channels through which they can seek it and identifying the entities from whom remedy can be sought. Supervisory authorities, including regulators, consumer protection authorities and judicial review entities, must possess sufficient resources and expertise to effectively monitor potential harms within their jurisdictions and intervene when required. While AI regulations are advancing globally, the absence of clear regulatory bodies for AI governance introduces ambiguity in responsibilities and enforcement mechanisms. This ambiguity poses a challenge to the authorities' proactive identification and resolution of emerging issues in the AI landscape.

Developers and deployers of AI: The UN Guiding Principles for access to remedy within technology sector recommend that businesses (in this case developers and deployers of AI) to "establish or participate in effective operational-level grievance mechanisms" (United Nations Human Rights Office of the High Commissioner, 2023b). Such mechanisms should be legitimate (i.e. enabling trust); accessible; predictable; equitable; transparent; rights-compatible; a source of continuous learning; and based on engagement and dialogue with stakeholders. This requires to implement well-designed remedies that respect the needs and objectives of affected individuals or groups. Patronizing and culturally insensitive communication should be avoided. Furthermore, regular collection and analysis of feedback on people's experiences in seeking remedies are important. This process helps identify

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

whether procedures are overly complicated and demand an excessively high level of digital literacy, promoting a more accessible and user-friendly redress mechanism.

One significant issue with current AI regulation efforts such as EU AI Act is their predominant focus on AI providers or general purpose AI models (Wachter, 2024), often overlooking the deployers who are closer to users and have a deeper understanding of the harms in their specific deployed context. However, the intricate relationship between developers and deployers in both technological and business aspects necessitates a collaborative system. Establishing an effective communication system is vital for highlighting the capabilities and limitations of AI and for incorporating feedback and complaints from affected individuals.

Beyond attributing liability to specific entities, another primary goal is to cultivate an environment where responsible AI development is a collective responsibility. This collaborative approach is essential for preventing reckless development, deployment and misuse of AI systems and for ensuring timely improvements and adjustments. AI providers must ensure clear communication with deployers regarding the system's capabilities, limitations and correct usage conditions, ensuring that AI systems are deployed in accordance with their design intentions. When deployers encounter malfunctions or issues stemming from the AI's design or functionality, they should be able to hold developers accountable for any resulting harm to their operations and affected individuals. Given their proximity to users and affected parties, deployers are uniquely positioned to detect and report specific harms or failures that occur during deployment. Therefore, it is crucial for deployers to have clear and effective mechanisms for providing redress and preventing the recurrence of similar issues. Implementing these collaborative approaches should focus on two key aspects: providing timely and adequate remedies for affected parties and creating a feedback loop for the continuous improvement of AI development and deployment.

Civil society and independent researchers: It has been acknowledged that seeking redress is often challenging, if not entirely impossible. Those challenges often stem from the technical complexity of AI, complexity in AI business relationships, regulatory hurdles and various legal, practical and procedural barriers (United Nations Human Rights Office of the High Commissioner, 2023c). Civil societies and independent researchers can play a vital role in empowering individuals to effectively navigate these intricacies surrounding redress mechanisms, especially in the realm of AI where affected individuals or communities may have a limited understanding of the complex services they are interacting with (Ogunleye, 2022). In recent years, civil society organizations conducting independent investigations or research have been instrumental in bringing significant public attention to most harmful or biased applications of AI systems (such as the Gender Shades project (Buolamwini and Gebru, 2018) and ProPublica's Machine Bias project (Julia et al., 2016)). Moreover, civil society organizations, such as consumer advocacy groups, civil right groups and other public interest entities, can play a crucial role in representing harmed groups and facilitating collective redress efforts. For example, in Cambridge Analytica scandal, digital rights groups like Electronic Frontier Foundation played a pivotal role in informing affected users (Gebhart, 2018) and supporting FTC's class-action lawsuits covering 250–280 million Facebook users to hold the company accountable (Raymond, 2022). By acting as representatives and facilitators for collective redress, these organizations not only amplify the voices of those who might otherwise go unheard but also provide the necessary resources and expertise to navigate complex redress systems, helping to ensure that collective harms are addressed effectively.

In this paper, we argue for the establishment of effective redress mechanisms as critical ex-post measures in AI governance that protect all individuals affected by AI. We highlight that the current landscape's overemphasis on ex-ante measures leaves impacted individuals without adequate means to rectify their situations. We outline four critical steps to seek redress, along with the challenges faced at each step. Acknowledging the interdisciplinary nature of developing and implementing these mechanisms, we highlight the need for increased collaboration and research across law, social sciences and data science. Access to redress is a fundamental component of access to justice, ensuring

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

that individuals have the opportunity to seek remedies for the wrongs they have suffered. By providing channels for individuals to effectively address harm caused by AI – whether through formal legal processes or informal pathways such as an ombudsman – individual affected can be empowered to hold accountable those responsible for harm. This access not only restores individual rights but also fosters a sense of fairness and trust in AI governance, reinforcing the principle that justice should be accessible to all.

Acknowledgements. Yulu thanks the Warwick Chancellor's Scholarship and the Turing Enrichment Fellowship for their generous support. Maddie was supported by an ERA Fellowship. We are also grateful to Alan Chan, Kerry McInerney, and Eleanor Drage for their valuable feedback on earlier drafts.

References

Ada Lovelace Institute. (2021). Keeping an eye on AI. Accessed: 2024-10-17. https://www.adalovelaceinstitute.org/report/keeping-an-eye-on-ai/.

Ahmed, S. (2021). Complaint as feminist pedagogy. https://feministkilljoys.com/2021/06/16/complaint-as-feminist-pedagogy/.

Alfrink, K., Keller, I., & Kortuem, G. (2022). Contestable AI by design: Towards a framework. *Minds & Machines*, *33*(2023), 613–639. https://doi.org/10.1007/s11023-022-09611-z.

Almada, M., & Petit, N. (2023). The EU AI act: A medley of product safety and fundamental rights? *Technical report 2023/59*. European University Institute (EUI). Retrieved from Cadmus, EUI Research Repository, https://hdl.handle.net/1814/75982.

Amarasinghe, K., Rodolfa, K. T., Jesus, S., Chen, V., Balayan, V., Saleiro, P., Bizarro, P., Talwalkar, A., & Ghani, R. (2024). On the importance of application-grounded experimental design for evaluating explainable ML methods. *Proceedings of the AAAI Conference on Artificial Intelligence*. *38*(19), 20921–20929. https://doi.org/10.1609/aaai.v38i19.30082.

Amsterdam District Court. (2021). Decision of the Amsterdam District Court. Accessed: 2024-08-14. https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2021:1019.

Board, E. D. P. (2021). Finnish SA: Police reprimanded for illegal processing of personal data with facial recognition software. https://www.edpb.europa.eu/news/national-news/2021/finnish-sa-police-reprimanded-illegal-processing-personal-data-facial_en.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R. E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S. M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., & Liang, P. (2021). On the opportunities and risks of foundation models, *CoRR*. abs/2108.07258. https://arxiv.org/abs/2108.07258.

Bommasani, R., Arora, S., Choi, Y., Ho, D.E., Jurafsky, D., Koyejo, S., Lakkaraju, H., Li, F.-F., Narayanan, A., Nelson, A., Pierson, E., Pineau, J., Varoquaux, G., Venkatasubramanian, S., Stoica, I., Liang, P., & Song, D. (2024). A path for science- and evidence-based AI policy. Accessed: 2024-09-27. https://understanding-ai-safety.org/.

Brookings. (2023). AI poses disproportionate risks to women. Accessed: 2024-09-15. https://www.brookings.edu/articles/ai-poses-disproportionate-risks-to-women/#:text=AI%2Dinduced%20job%20losses%20are,policies%20to%20promote%20gender%20equity.

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. The MIT Press.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., héigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. https://arxiv.org/abs/1802.07228.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. https://proceedings.mlr.press/v81/buolamwini18a.html.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms, *Big Data & Society*. *3*(1), https://doi.org/10.1177/2053951715622512.

Business and Human Rights Resource Centre. (2016). Remediation and grievance mechanisms. https://www.businessrespecthumanrights.org/en/page/349/remediation-and-grievance-mechanisms.

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, *81*(5), 825–836, https://misclibrary.wiley.com/doi/abs/10.1111/puar.13293 (eprint: https://misclibrary.wiley.com/doi/pdf/10.1111/puar.13293)

Cobbe, J., Veale, M., & Singh, J. (2023a). Understanding accountability in algorithmic supply chains. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT'23), Chicago, IL, United States. ACM, pp. 1–12. https://doi.org/10.1145/3593013.3594073.

Cobbe, J., Veale, M., & Singh, J. (2023b). Understanding accountability in algorithmic supply chains. 2023 ACM Conference on Fairness, Accountability, and Transparency, pp. 1186–1197, ACM, Chicago, IL, United States. https://dl.acm.org/doi/10.1145/3593013.3594073.

Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8.

Committee, T. N. A. A. (2023). Recommendation: Improve monitoring of emerging risks from AI through adverse event reporting. Accessed: 2024-09-27. https://ai.gov/wp-content/uploads/2023/12/Recommendation_Improve-Monitoring-of-Emerging-Risks-from-AI-through-Adverse-Event-Reporting.pdf.

Competition and Markets Authority. (2024). CMA annual plan 2024 to 2025. https://www.gov.uk/government/publications/cma-annual-plan-2024-to-2025/annual-plan-2024-to-2025.

Davies, M., & Birtwistle, M. (2023). Regulating AI in the UK. https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/.

Demková, S. (2023). Automated decision-making and effective remedies, Edward Elgar Publishing. https://www.elgaronline.com/monochap/book/9781035306619/book-part-9781035306619-6.xml.

DHH. (2019). Accessed: 2024-08-14. https://x.com/dhh/status/1192540900393705474.

Digital Regulation Platform. (2020). Dispute resolution and redress. Last updated on: 19.01.2022. https://digitalregulation.org/dispute-resolution-and-redress/.

DiMatteo, L. A., Poncibò, C., & Cannarsa, M. (2022). AI and liability. *The Cambridge handbook of artificial intelligence: Global perspectives on law and ethics* (pp. 87–160). Cambridge, Cambridge Law Handbooks, Cambridge University Press.

Drage, E., McInerney, K., & Browne, J. (2024). Engineers on responsibility: Feminist approaches to who's responsible for ethical AI. *Ethics and Information Technology*, *26*(1), 4, https://doi.org/10.1007/s10676-023-09739-1.

Dunlop, C. (2023). An EU AI act that works for people and society. https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act-trilogues/#:text=5.-,Protection%20and%20representation%20for%20affected%20persons,proposed%20by%20the%20European%20Parliament./.

Ehsan, U., Singh, R., Metcalf, J., & Riedl, M. (2022). The algorithmic imprint. 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT'22. ACM, http://dx.doi.org/10.1145/3531146.3533186.

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, *5*, 40–60. https://estsjournal.org/index.php/ests/article/view/260.

Engler, A. (2023a). The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment. Accessed: 2025-01-23. https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/.

Engler, A. (2023b). Key enforcement issues of the AI act should lead EU trilogue debate. https://www.brookings.edu/articles/key-enforcement-issues-of-the-ai-act-should-lead-eu-trilogue-debate/.

EU AI Act. (2024). Article 86: Right to explanation of individual decision-making. Accessed: 2024-08-20. https://artificialintelligenceact.eu/article/86/#:text=This%20article%20states%20that%20if,in%20the%20decision%2Dmaking%20process.

European Commission. (2024a). Annex 1 of the artificial intelligence act. https://artificialintelligenceact.eu/annex/1/.

European Commission. (2024b). Representative actions directive. https://commission.europa.eu/law/law-topic/consumer-protection-law/representative-actions-directive_en.

European Commission. (2025). 2025 commission work programme and annexes. https://commission.europa.eu/publications/2025-commission-work-programme-and-annexes_en.

European Convention on Human Rights. (1950). Article 13 of the convention – Right to an effective remedy. Accessed: 2024-09-15. https://www.echr.coe.int/Documents/Convention_ENG.pdf.

European Council. (2024). Eu brings product liability rules in line with digital age and circular economy. https://www.consilium.europa.eu/en/press/press-releases/2024/10/10/eu-brings-product-liability-rules-in-line-with-digital-age-and-circular-economy/.

European Parliament. (2023). EU legislation in progress: Artificial intelligence liability directive. Accessed: 2024-08-20. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf.

European Parliament. (2025). Revised product liability directive. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739341/EPRS_BRI(2023)739341_EN.pdf.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Fanni, R., Steinkogler, V. E., Zampedri, G., & Pierson, J. (2023). Enhancing human agency through redress in Artificial Intelligence Systems. *AI & SOCIETY*, *38*(2), 537–547. https://doi.org/10.1007/s00146-022-01454-7.

Federal Trade Commission. (2021). FTC refund programs. https://www.ftc.gov/enforcement/refunds.

Federal Trade Commission. (2023). Ftc law enforcers nationwide announce enforcement sweep to stem tide of illegal telemarketing calls in the U.S. https://www.ftc.gov/news-events/news/press-releases/2023/07/ftc-law-enforcers-nationwide-announce-enforcement-sweep-stem-tide-illegal-telemarketing-calls-us.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, *6*(1), 2053951719860542. https://doi.org/10.1177/2053951719860542.

Ferguson, A. N., Franklin, M., & Lagnado, D. (2022). Explanations that backfire: Explainable artificial intelligence can cause information overload. Proceedings of the Annual Meeting of the Cognitive Science Society.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707, Epub 2018 Nov 26. PMID: 30930541; PMCID: PMC6404626.

Gallen, J., & Moffett, L. (2022). The palliative role of reparations in reconciling societies with the past: Redressing victims or consolidating the state? *Journal of Intervention and Statebuilding*. *16*(4), 498–518. https://doi.org/10.1080/17502977.2022.2042650.

Gebhart, G. (2018). How to change your Facebook settings to opt out of platform API sharing. Accessed: 2025-01-23. https://www.eff.org/deeplinks/2018/03/how-change-your-facebook-settings-opt-out-platform-api-sharing.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(2015), 452–459. https://doi.org/10.1038/nature14541.

Goertzel, K. (2016). Legal liability for bad software. *CrossTalk*, *29*(September/October, 2016), 23–28. https://zlmonroe.com/CSE566/Readings/47.Legal_Liability_for_Bad_Software.pdf.

Government of the United Kingdom. (2021). Tackling violence against women and girls. https://assets.publishing.service.gov.uk/media/6194d05bd3bf7f054f43e011/Tackling_Violence_Against_Women_and_Girls_Strategy_-_July_2021.pdf.

Government of United Kingdom. (2025). Complaints procedure. https://www.gov.uk/government/organisations/monitor/about/complaints-procedure.

Greene, T. (2023). US surveillance and facial recognition firm Clearview AI wins GDPR appeal in UK court. https://cointelegraph.com/authors/tristan-greene.

Hacker, P. (2023). The European AI liability directives – Critique of a half-hearted approach and lessons for the future, *Computer Law & Security Review*, *51*(2023), 105871. https://doi.org/10.1016/j.clsr.2023.105871. https://ssrn.com/abstract=4279796.

Henley, J. (2021). Dutch government resigns over child benefits scandal. Accessed: 2024-09-23. https://www.theguardian.com/world/2021/jan/15/dutch-government-resigns-over-child-benefits-scandal.

Hern, A. (2020). Twitter apologises for racist image-cropping algorithm. Accessed: 2024-09-23. https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm.

Hodges, C. (2014). Consumer redress: Ideology and empiricism. In K. Purnhagen, & P. Rott (Eds.), *Studies in European economic law and regulation* (Vol. 3, pp. 793–821). Springer. https://doi.org/10.1007/978-3-319-04903-8_39.

House, T. W. (2023). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

International Organization for Standardization. (2021). ISO/IEC TR 24027:2021 information technology—Artificial intelligence (AI)—Bias in AI systems and AI-aided decision making. Accessed: 2024-08-23. https://www.iso.org/standard/81230.html.

Johnson, K. (2023). Algorithms allegedly penalized black renters. The US government is watching. https://www.wired.com/story/algorithms-allegedly-penalized-black-renters-the-us-government-is-watching/.

Johnson, N., Moharana, S., Harrington, C., Andalibi, N., Heidari, H., & Eslami, M. (2024). The fall of an algorithm: Characterizing the dynamics toward abandonment. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, pp. 337–358, FAccT'24, Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3630106.3658910.

Jones, K. (2023). AI governance and human rights: Resetting the relationship. https://www.chathamhouse.org/sites/default/files/2023-01/2023-01-10-AI-governance-human-rights-jones.pdf.

Julia, A., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals and it's biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessmentsin-criminal-sentencing.

Khosrowi, D., Finn, F., & Clark, E. (2024). Engaging the many-hands problem of generative-AI outputs: A framework for attributing credit. *AI Ethics*. https://doi.org/10.1007/s43681-024-00440-7.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Kluttz, D. N., Kohli, N., & Mulligan, D. K. (2020). *Shaping our tools: Contestability as a means to promote responsible algorithmic decision making in the professions* (pp. 137–152). Cambridge University Press.

Kolt, N., Anderljung, M., Barnhart, J., Brass, A., Esvelt, K., Hadfield, G. K., Heim, L., Rodriguez, M., Sandbrink, J. B., & Woodside, T. (2024). Responsible reporting for frontier AI development. https://arxiv.org/abs/2404.02675.

Lutz, R. (2017). Incident number 96: Houston Schools Must Face Teacher Evaluation Lawsuit. In McGregor, S. (ed.), *Artificial Intelligence Incident Database. Responsible AI Collaborative*. Retrieved on 2025-04-24. https://incidentdatabase.ai/cite/96.

Lyons, H., Velloso, E., & Miller, T. (2021). Conceptualising contestability: Perspectives on contesting algorithmic decisions. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. http://dx.doi.org/10.1145/3449180.

Mansi, G., & Riedl, M. (2023). Why don't you do something about it? Outlining connections between AI explanations and user actions.

Metcalf, J., Singh, R., Moss, E., Tafesse, E., & Watkins, E. A. (2023). Taking algorithms to courts: A relational approach to algorithmic accountability. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pp. 1450–1462, FAccT'23, Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3593013.3594092.

Mitchell, T. (2024). Google's AI Gemini is generating inaccurate historical information. Accessed: 2024-09-23. https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical.

National Institute of Standards and Technology. (2023). AI risk management framework. Accessed: 2024-08-23. https://www.nist.gov/itl/ai-risk-management-framework.

New York State Department of Financial Services. (2021). Apple card investigation report. Accessed: 2024-08-14. https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://www.science.org/doi/abs/10.1126/science.aax2342.

Ogunleye, I. (2022). *Recommendations to improve consumer protection from artificial intelligence*. https://cltc.berkeley.edu/wp-content/uploads/2022/08/AIs_Redress_Problem.pdf.

Oh, C., Song, J., Choi, J., Kim, S., Lee, S., & Suh, B. (2018). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–13, CHI'18, Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3173574.3174223.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, Crown Publishing Group.

Organisation for Economic Co-operation and Development. (2019). OECD AI principles. Accessed: 2024-08-23. https://oecd.ai/en/ai-principles.

Ploug, T., & Holm, S. (2020). The four dimensions of contestable AI diagnostics – A patient-centric approach to explainable AI. *Artificial Intelligence in Medicine*, *107*(July 2020), 101901. https://www.sciencedirect.com/science/article/pii/S0933365720301330.

Raji, I. D., Kumar, I. E., Horowitz, A., & Selbst, A. (2022). The fallacy of AI functionality. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 959–972, FAccT'22, Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3531146.3533158.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *CoRR*, http://arxiv.org/abs/2001.00973.

Raymond, N. (2022). Facebook parent Meta to settle Cambridge Analytica scandal case for $725 million. Accessed: 2025-01-23. https://www.reuters.com/legal/facebook-parent-meta-pay-725-mln-settle-lawsuit-relating-cambridge-analytica-2022-12-23/.

Redress. (2024a). The importance of legal redress. Accessed: 2024-08-23. https://redress.org/news/the-importance-of-legal-redress/.

REDRESS. (2024b). Training module 15: Evidence of torture. Accessed: 2024-08-20. https://redress.org/wp-content/uploads/2024/06/Training-Module-15-Evidence-Of-Torture.pdf.

Rhodes, C., Rough, E., Panjwani, A., Booth, L., Alston, G., Little, P., & Robertson, T. (2024). *Debate on access to redress schemes*. House of Commons Library, Research Briefing. https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0077/.

Rights, E. D. (2023). An EU artificial intelligence act for fundamental rights a civil society statement. https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf.

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction*, *35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391.

Scottish Public Services Ombudsman. (2017). Redress policy. Accessed: 2024-08-23. https://www.spso.org.uk/sites/spso/files/communications_material/RedressPolicy.pdf.

https://doi.org/10.1017/cfl.2025.9 Published online by Cambridge University Press

Sejal Goud, Aaron F. Mertz, Jylana L. Sheats, Dorothy Chou (2023). A blueprint for equitable AI, *Technical report*, The Aspen Institute. https://www.aspeninstitute.org/wp-content/uploads/2023/01/Equitable-AI-Aspen-Institute.pdf.

Selander, L., Jarvenpaa, S., & Kronblad, C. (2023). Awakening to algorithmic transgressions: Non-users discovery of algorithmic decision making, *Academy of Management Proceedings*, *2023*(1), 19344. https://doi.org/10.5465/AMPROC.2023.17bp.

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, *7*(4), 233–242. https://doi.org/10.1093/idpl/ipx022.

Smuha, N. A. (2021). Beyond the individual: Governing AI's societal harm. *Internet Policy Review*, *10*(3), https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm.

Stefan Zagelmeyer, L. B., & Shemberg, A. R. (2018). Non-state based non-judicial grievance mechanisms (NSBGM): An exploratory analysis. Accessed: 2024-08-14. https://www.ohchr.org/sites/default/files/Documents/Issues/Business/ARP/ManchesterStudy.pdf.

Stockhem, O. (2023). Access to justice and effective remedy in the EU AI act: The state of play. https://cdt.org/insights/access-to-justice-and-effective-remedy-in-the-eu-ai-act-the-state-of-play/.

Tréhu, J., & Ricart, R. J. (2024). Global AI governance: Key steps for transatlantic cooperation. Accessed: 2025-01-23. https://www.gmfus.org/sites/default/files/2024-11/ECA%20AI%20POLICY%20REPORT%20VER%206%5B46%5D.pdf.

UK Department for Science, I. and Technology. (2023). A pro-innovation approach to AI regulation. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.

UK Finance. (2021). Reforming statutory dispute resolution processes in banking and finance. https://www.ukfinance.org.uk/news-and-insight/blogs/reforming-statutory-dispute-resolution-processes-banking-and-finance.

UK's Department for Science, I. and Technology. (2024). Implementing the UK's AI regulatory principles: Initial guidance for regulators. https://assets.publishing.service.gov.uk/media/65c0b6bd63a23d0013c821a0/implementing_the_uk_ai_regulatory_principles_guidance_for_regulators.pdf.

Union, E. (2021). EU AI act: First regulation on artificial intelligence. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

United Nations. (1948). Universal declaration of human rights. https://www.un.org/en/udhrbook/pdf/udhr_booklet_en_web.pdf.

United Nations Educational, S., & Organization, C. (2021). Recommendation on the ethics of artificial intelligence. Accessed: 2024-08-23. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

United Nations Human Rights Office of the High Commissioner. (2023a). Access to remedy and the technology sector: A "remedy ecosystem" approach. https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-ecosystem-approach.pdf.

United Nations Human Rights Office of the High Commissioner. (2023b). Access to remedy and the technology sector: Basic concepts and principles. https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf.

United Nations Human Rights Office of the High Commissioner. (2023c). Access to remedy and the technology sector: Understanding the perspectives and needs of affected people and groups. https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-perspectives-needs-affected-people.pdf.

U.S. Equal Employment Opportunity Commission. (2024). iTutorGroup to pay 365,000 dollar to settle EEOC discriminatory hiring suit. https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit#:text=According%20to%20the%20EEOC's%20lawsuit,States%20because%20of%20their%20age.

Vickers, M. (2022). Civil justice reform: An ombudsman perspective. *Amicus Curiae*, *4*(1). https://journals.sas.ac.uk/amicus/article/view/5494.

Wachter, S. (2024). 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 Journal of Law and Technology*, *26*(3), 671–718.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76–99, https://doi.org/10.1093/idpl/ipx005.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent abilities of large language models. https://arxiv.org/abs/2206.07682.

Wendehorst, C. (2020). Strict liability for AI and other emerging technologies. *Journal of European Tort Law*, *11*(2), 150–180. https://www.degruyter.com/document/doi/10.1515/jetl-2020-0140/html?lang=en.

Cite this article: Pi Y. and Proctor M. (2025). Toward empowering AI governance with redress mechanisms. *Cambridge Forum on AI: Law and Governance* **1**, e24, 1–22. https://doi.org/10.1017/cfl.2025.9