

Information Privacy in the Age of AI: A Narrative Review of Enforceability and Privacy by Design under the UK GDPR

Wala Mohammed¹

MSc Management of Information Systems and Digital Innovation
Department of Management
London School of Economics and Political Science

KEYWORDS	ABSTRACT
Information Privacy, Generative AI, Automated Decision-Making System, UK GDPR, Privacy by Design, Algorithmic Bias, Sociotechnical Perspective, Socio-Technical Misalignment.	This narrative review examines whether the United Kingdom's data protection regime can withstand the technical realities of artificial intelligence (AI). As Generative AI (GenAI) and Automated Decision-Making systems (ADMs) become embedded across the public and private sectors, the study addresses two analytically connected questions, focused on whether the data subject rights at the core of the United Kingdom (UK) GDPR remain practically enforceable in the age of AI, and whether Privacy by Design constitutes an achievable technical reality for these systems. The paper adopts a critical, narrative literature review, grounded in the Traditional Sociotechnical Perspective. Literature identified from sources published between 2006 and 2026, through structured searches of Scopus, Web of Science, the AIS eLibrary, and Google Scholar, supplemented by primary legislation and official regulatory guidance from GOV.UK and the Information Commissioner's Office. The review finds that architectural features of deep learning models, particularly the encoding of training data in model weights and the opacity of model reasoning, render the rights to erasure, rectification, and explanation practically unenforceable. It argues that the Data (Use and Access) Act 2025 widens this gap, and that privacy-preserving technologies offer only partial, flawed remedies. Its contribution lies in reframing enforceability as socio-technical misalignment and proposing a layered programme of legal, institutional, and design reform. Limitation particularly those concerning the recent Data (Use and Access) Act 2025, rest substantially on scholarly interpretation rather than established empirical evidence or case law, and should be read as anticipatory.

¹Wala Mohammed
Email Address: w.a.mohammed@lse.ac.uk

1. Introduction

Privacy has long been recognised as a fundamental human right, yet its definition remains contested across legal, cultural, and philosophical traditions (Barth et al., 2022). Westin's foundational conception frames privacy as the right of individuals to determine when, how, and to what extent information about them is shared with others (Barth et al., 2022). Solove situates this debate in a broader global context, showing that the challenge of defining privacy transcends national boundaries and is reflected in constitutions and international legal instruments across diverse traditions (Solove, 2006). Hence, privacy is not merely an academic concern but a heavily legislated global issue facing unprecedented strain (Solove, 2006). The proliferation of Artificial Intelligence (AI) across both public and private sectors has fundamentally transformed how personal data is collected and processed (Feretzakis et al., 2025; Janssen, 2020; Sargeant, 2026). Article 4(1) of the General Data Protection Regulation (GDPR) operationalises this concept by defining personal data as “any information relating to an identified or identifiable natural person” (Ohm, 2009).

This literature review focuses on information privacy in two systems, Generative AI (GenAI) and Automated Decision-Making Systems (ADMs) (Duffourc et al., 2024; Marabelli et al., 2021). Information Privacy is categorised as one of the four main dimensions of privacy alongside communication, territory, and the body, and deals explicitly with the lifecycle of personal data from multiple aspects, including how data is aggregated, stored, processed, anonymised, shared, used, and discarded (Majeed & Hwang, 2023). While existing literature has extensively examined the legal provisions of the GDPR and the technical capabilities of privacy-preserving technologies in isolation, there remains a significant gap in research that critically evaluates the practical enforceability of data

protection rights in the UK against the specific architectural design of GenAI and ADM systems operating in both public and private sector contexts (Feretzakis et al., 2025; Janssen, 2020; Sargeant, 2026). Furthermore, despite growing academic attention to Privacy by Design as a regulatory principle, there is insufficient analysis of whether it constitutes an achievable technical reality in the development of both GenAI and ADMs (Drake et al., 2022; Duffourc et al., 2024; Feretzakis et al., 2025; Kingsman et al., 2022; Koops & Leenes, 2014; Sargeant, 2026).

Given this background, this literature review addresses two analytically connected research questions. First, whether the UK GDPR remains effective at protecting information privacy in the age of AI (RQ1); and second, whether Privacy by Design constitutes an achievable technical reality in the context of GenAI and ADMs (RQ2). The two questions are connected sequentially and evaluatively. RQ1 diagnoses the problem, assessing the extent to which the enforceability of data subject rights survives contact with the architecture of these systems. RQ2 then interrogates the law's own principal remedial mechanism; Article 25 of the UK GDPR presumes that the rights examined under RQ1 can be secured proactively through technical and organisational design. If Privacy by Design proves technically unachievable for GenAI and ADMs, the enforceability gap identified under RQ1 cannot be closed from within the existing framework, and the case for external legal and institutional reform follows directly. The two questions, therefore, jointly test whether the current data protection regime is capable of self-correction. To keep this scope tractable, the review treats the enforceability of data subject rights against deep learning architectures as its central analytical problem; it examines adjacent issues, including data brokerage, sectoral boundaries, and consent, only insofar as they condition that problem. By synthesising current scholarship, this review argues that existing frameworks are fundamentally misaligned with

the technical realities they seek to govern, and that urgent reform, spanning legal, institutional, and design dimensions, is required to close the widening gap between regulatory intent and practical protection (Drake et al., 2022; Duffourc et al., 2024; Feretzakis et al., 2025; Kingsman et al., 2022; Koops & Leenes, 2014; Sargeant, 2026). This literature review is structured into four sections; Section 2 provides a contextual background on the conceptualisation of privacy and the divergent adoption of GenAI and ADMs across the public and private sectors; Section 3 presents the findings, examining the failure of data subject rights in the AI age, the dismantling of automated decision-making protections, the erosion of privacy across blurred sectoral boundaries, and the promise and limitations of privacy-preserving technologies. Finally, Section 4 synthesises these findings into a discussion of recommendations towards a more effective and enforceable data protection framework.

2. Conceptual Framework

The literature conceptualises information privacy through several distinct lenses, ranging from moral imperatives to economic transactions (Smith et al., 2011). At one end of the spectrum, the fundamental human rights perspective views privacy as a moral and legal necessity embedded within the fabric of society (Smith et al., 2011). Conversely, the commodity view treats privacy through an economic lens, where individuals approach their personal data as an asset to be traded or surrendered based on a calculated cost-benefit analysis, also known as privacy calculus (Smith et al., 2011). An alternative perspective defines privacy through the lens of cognitive ability, shifting the focus to human cognition to frame privacy as a dynamic interaction between access and agency (Smith et al., 2011). In this view, privacy is characterised as a state of limited access, emphasising an individual's proactive ability to selectively regulate who can reach their personal data, allowing them to strategically

manage and minimize their own vulnerability (Smith et al., 2011).

Technology has raised individual online privacy concerns (OPC), which refer to the fear of unauthorized data collection and manipulation, prompting consumers to adopt defensive behaviors, such as boycotting platforms, using ad blockers, or fabricating personal details (Anic et al., 2019). Whilst weak regulations and a heightened sense of privacy entitlement intensify individuals' online privacy concerns, high levels of institutional trust and the perceived benefits of digital services can significantly mitigate these anxieties (Anic et al., 2019). Indeed, the convenience and productivity gains associated with AI-powered services frequently outweigh privacy considerations, leading users to prioritise access to digital services over protecting their personal data (Akakpo et al., 2025; Anic et al., 2019). Early experiments demonstrated that users who expressed strong privacy concerns have willingly engaged in personal dialogues with automated bots and answered highly sensitive questions (Kokolakis, 2017).

Another central tension identified across the literature is the gap between the legal ideal of informed consent and its practical realisation (Andreotta et al., 2022). Under the GDPR, valid consent must be freely given, specific, informed, and unambiguous, and individuals must retain the right to withdraw it at any time (Whitley & Pujadas, 2018). In practice, digital platforms force users to either accept all terms and conditions (T&Cs) or lose access entirely (Andreotta et al., 2022). Many users in the UK do not read privacy notices or T&Cs, while many service providers remain deliberately vague about critical technical details such as data retention periods and security protocols (Whitley & Pujadas, 2018). T&Cs have proven consistently ineffective at communicating data processing activities to ordinary users, given consumers' limited capacity to process complex legal and technical information, which has been

identified as a central barrier to meaningful engagement (Whitley & Pujadas, 2018).

The rise of digital platforms has fundamentally transformed personal data into a tradeable commodity of significant economic value (Antonio, 2006; Bonina et al., 2021; Murray, 2010). The ability to trade data across platforms and devices has given rise to a multi-billion-pound industry (Murray, 2010). The private sector has institutionalised the commercial exploitation of personal data through data brokers, who trade tailored consumer insights to facilitate targeted advertising on these platforms (D'Hauwers, 2026; Murray, 2010). Data brokers are commercial entities that collect personal data from different sources, process it into detailed consumer profiles, and monetise it by offering informational goods and services to third parties, typically without the direct control or knowledge of the individuals (DHauwers, 2026). Within this ecosystem, the highest-risk actors are those collaborating across shared technical environments to trade personal data, whilst AI platforms pose an acute threat through the systematic scraping of personal data to train algorithms and construct detailed user profiles (DHauwers, 2026).

2.1 The Deployment of GenAI and ADMs Across Public and Private Sectors

The private and public sectors follow divergent trajectories in adopting machine learning models (Wang et al., 2025). Private-sector actors have driven the development of advanced machine learning systems, shaped by profit-driven imperatives rather than broader social objectives (Margetts & Dorobantu, 2019). The public sector, however, has fallen behind the private sector in technological capacity and development (MacArthur et al., 2022). From the 1980s, governments shifted away from resilience, instead prioritising outsourcing over developing their own infrastructure, and using technology simply to cut costs and automate

routine tasks (MacArthur et al., 2022). Governments have increasingly turned to these technologies in response to their well-documented economic advantages, notably greater cost-effectiveness and productivity gains (Wang et al., 2025). When AI has been deployed in the UK public sector, it has frequently exposed rather than resolved pre-existing social inequalities (Margetts & Dorobantu, 2019).

Both the private and public sectors have adopted several deep learning applications (Aldemir & Uysal, 2025; Duffourc et al., 2024; Kshetri, 2025; Sargeant, 2026; Suksi, 2021). For instance, Generative Artificial Intelligence (GenAI) refers to AI models that generate content like text, images, videos, and sounds from user prompts (Duffourc et al., 2024). Users who seek personalised interactions with these data-rich models face an inherent trade-off, as generating tailored AI content requires ongoing disclosure of personal data (Lim & Shim, 2022; Meng & Liu, 2025). For instance, integrating GenAI assistants into web browsers has also intensified privacy concerns beyond those posed by traditional digital platforms (Vekaria et al., 2025). Unlike conventional browser extensions, GenAI extensions are fundamentally more invasive due to their architectural privileges and the way they process information (Vekaria et al., 2025). To deliver personalised responses, these extensions constantly monitor user activity and rely on server-side processing, meaning that user queries and contextual data are transmitted entirely off the local device to proprietary models, effectively enabling continuous surveillance (Vekaria et al., 2025).

ADMs, on the other hand, are defined as organisational socio-technical systems that automate or support strategic decisions, with consequences within and beyond organisational boundaries (Marabelli et al., 2021). ADMs introduce severe risks; historical bias in data can lead to discriminatory outputs (Janssen, 2020).

Algorithmic bias in ADMs often arises when the datasets used to train these models fail to adequately represent the full diversity of the population (Vogelsang & Borg, 2019). In certain applications, the inclusion of demographic data can introduce discriminatory outcomes, for instance, by unfairly restricting individuals' access to essential services if race or gender is included in the training dataset (Vogelsang & Borg, 2019). Private companies in the UK are increasingly deploying ADMs for targeted advertising, credit scoring, and healthcare services (Janssen, 2020). The public sector is integrating ADMs to reduce delays in processing public service requests and better meet rising public expectations (Aldemir & Uysal, 2025). In 2020, for example, an algorithm deployed to grade students automatically and unfairly downgraded students from poorer areas based on historical school performance, eroding public trust (Marabelli et al., 2021). The UK Government has continued to adopt these modern technologies on the basis that the economic benefits outweigh the risks (Sargeant, 2026).

2.2 Review Methodology and Limitations

This paper takes the form of a critical, narrative literature review, an approach suited to synthesising and problematising a heterogeneous body of legal, technical, and Information Systems scholarship around a focal argument (Paré et al., 2015; Webster & Watson, 2002). Literature was identified through structured searches of Scopus, Web of Science, the AIS eLibrary, and Google Scholar, complemented by searches of official legal sources, including GOV.UK, and the Information Commissioner's Office (ICO) guidance repository. Search terms combined privacy-related keywords ('information privacy', 'data protection', 'GDPR', 'privacy by design', 'consent') with technology-related keywords ('generative AI', 'large language

models', 'automated decision-making', 'machine unlearning', 'federated learning', 'differential privacy'). Backward and forward citation chaining from key contributions was used to identify additional sources (Webster & Watson, 2002).

Sources were included if they were published in English between 2006 and 2026 and addressed the intersection of data protection law and the technical architecture of GenAI or ADMs; purely doctrinal analyses of unrelated GDPR provisions and purely technical papers without privacy implications were excluded. Reflecting the material's heterogeneity, sources were classified and weighted by type. Peer-reviewed academic literature was treated as the primary evidential base; primary legislation and official regulatory guidance (GOV.UK, 2025; ICO, 2026) were used to establish the current state of the law; and legal practitioner commentary and preprints were drawn upon only for very recent developments not yet addressed in peer-reviewed work, and are identified as such in the text. The analysis then proceeded thematically; sources were coded against the two research questions and the organising concept of socio-technical misalignment, grounded in the Traditional Sociotechnical Perspective (Sarker et al., 2019), and synthesised into the four thematic findings presented in Section 3.

This review has several limitations that qualify its findings. Its synthesis reflects an interpretive selection and weighting of sources around a focal argument, and does not claim the exhaustive reproducibility of a systematic protocol (Paré et al., 2015). More consequentially, the intersection of deep learning architecture and data protection law remains a nascent field, and the pace of technological and regulatory change has outpaced the production of peer-reviewed empirical research. Several of the review's central claims, particularly those concerning the Data (Use and Access) Act 2025, therefore rest substantially on

scholarly interpretation, legal practitioner commentary, and preprints rather than on established empirical study, case law, or regulatory enforcement, none of which yet exist for a statute this recent. This reliance is a necessary consequence of examining a live and rapidly evolving area, but it means that a number of the evaluative conclusions advanced here are anticipatory in character, tracing the probable direction of a regulatory gap rather than documenting a settled one. As the field matures, empirical evidence of how these provisions operate in practice, alongside emerging case law and ICO enforcement, may confirm, qualify, or unsettle the arguments synthesised in this review. Its contribution should accordingly be read as a timely diagnostic framing of an unresolved problem rather than a definitive account, and as a foundation for the empirical and doctrinal research the field now urgently requires.

3. Findings: The Deterioration of Privacy Rights in the Digital Age

The rapid evolution of AI technologies has consistently outpaced the creation of new legislation, leaving only about 32% of the UK public trusting the AI sector (Drake et al., 2022). The UK and EU GDPR share broadly similar provisions, centered on giving individuals meaningful control over their personal data (Feretzakis et al., 2025). However, the opaque, data-hungry architectural design of modern deep learning models has introduced fundamental tension to the strict legal requirements of the GDPR (Feretzakis et al., 2025). Both GenAI and ADM systems are often underpinned by the same deep learning infrastructure, utilising neural network and transformer architectures pre-trained on vast and diverse datasets, producing outputs ranging from human-like text to consequential automated decisions (Feretzakis et al., 2025; Janssen, 2020; Sargeant, 2026). It is precisely this shared technical foundation that renders several core GDPR

provisions practically unenforceable across both technologies (Feretzakis et al., 2025; Janssen, 2020; Sargeant, 2026).

The post-Brexit regulatory landscape has evolved considerably, while the EU has addressed the limitations of the GDPR in governing algorithmic systems through enacting the EU AI Act (Drake et al., 2022; Feretzakis et al., 2025). However, the UK has favoured a softer regulatory approach, introducing the Algorithmic Transparency Standards while substituting Article 22 of its retained GDPR framework with the Data Use and Access Act 2025 (DUAA) (Kingsman et al., 2022; Sargeant, 2026). Among other countries, the UK has played a role in shaping ethical AI guidelines for the responsible development of these systems, with these frameworks converging around foundational interconnected principles, including privacy, fairness, and transparency (Jobin et al., 2019). Fairness requires that these systems are designed in a manner that does not perpetuate societal biases; privacy ensures sensitive information, such as race, is handled lawfully and with consent; and transparency binds both principles together by compelling organisations to openly disclose their data practices and the outcomes these systems generate (Jobin et al., 2019).

3.1 The Failure of Data Subject Rights in the AI Age

Scholars have argued that several core GDPR rights have become practically unenforceable because of these models' architectural design (Feretzakis et al., 2025; Sargeant, 2026). The Right to be Forgotten under Article 17 and the right to rectification under Article 16 allow individuals to request the deletion or correction of their personal data (Feretzakis et al., 2025). However, neither GenAI nor ADMs store data in traditional databases; instead, data is deeply encoded in neural network weights, making it technically infeasible to remove or edit specific data points without retraining the entire model

(Feretzakis et al., 2025; Janssen, 2020). Once data enters the GenAI lifecycle, users effectively lose their legal rights to correct or delete it (Duffourc et al., 2024; Feretzakis et al., 2025; Janssen, 2020). Furthermore, in GenAI, every prompt and user input submitted is retained and used by developers, including OpenAI and Google, as training data to enhance model performance, a practice that poses serious privacy risks (Duffourc et al., 2024). GenAI models pose a significant privacy threat through memorisation and training-data leakage, with experiments successfully extracting personal information directly from ChatGPT 3.5 (Duffourc et al., 2024; Feretzakis et al., 2025; Lermen et al., 2026). Scholars found that both Google and OpenAI guarantee they will delete user data upon request, yet they noted ambiguity about whether this data can truly be unlearned from core model parameters (Feretzakis et al., 2025). Furthermore, lawmakers have relied on anonymisation to balance privacy protection with the free flow of data and commerce (Ohm, 2009). While deanonymising online users is theoretically possible, it is too costly and labor-intensive to execute (Lermen et al., 2026). However, GenAI shifts this by drastically reducing the cost of operating at internet scale on raw, unstructured text without requiring manual investigation (Lermen et al., 2026). In a previous study, scholars found that by simply prompting an LLM with a summary of a pseudonymous profile is enough for the model to cross-reference details, search the web, and verify real-world identities in minutes, a task that would take a human investigator hours (Lermen et al., 2026).

3.2 *The Dismantling of Automated Decision-Making Protections*

The opacity of ADMs also directly undermines individuals' right to explanation (Feretzakis et al., 2025; Sargeant, 2026). Explainability is defined as the ability to understand the reasoning behind a specific decision; therefore, it is

distinct from mere transparency about a system's code or data and is far harder to achieve in practice (Sargeant, 2026). Under Article 15(1)(h), alongside Recital 71, individuals are entitled to meaningful information about the logic behind automated decisions in order to effectively contest them under Article 22 (Sargeant, 2026; UK GDPR, 2021). In practice, however, when an individual does not understand which input variables led to a negative outcome, such as a rejected credit application, they cannot effectively scrutinise the decision for discrimination or identify which inaccurate data requires correction (Feretzakis et al., 2025; Sargeant, 2026). This problem is compounded when two separate ADM systems interact, as one system's output becomes another's input, creating cascading chains of automated decisions whose reasoning becomes increasingly obscure and difficult to trace (Suksi, 2021).

The Data Use and Access Act 2025 (DUAA) has reshaped this landscape, and its consequences must be described with care, distinguishing legal description from evaluation (Kingsman et al., 2022; Sargeant, 2026). As a matter of description, the Act does not replace the UK GDPR; it amends it, replacing the former Article 22 restrictions with a more permissive framework under which significant decisions may be taken by solely automated means in wider circumstances, subject to statutory safeguards that include providing individuals with information about significant decisions, enabling them to make representations, human intervention, and to contest such decisions, with stricter conditions retained for special category data (GOV.UK, 2025; ICO, 2026). ICO guidance likewise continues to describe safeguards around automated decision-making, including information provision, human intervention, challenge mechanisms, and regular checks of systems for accuracy and bias (ICO, 2026). The evaluative claim advanced by several scholars, and endorsed with qualification in this review, is that this rebalancing weakens protection in practice; the statutory safeguards

presuppose precisely the explainability that the architectural opacity documented above undermines, since individuals cannot meaningfully contest, and human reviewers cannot meaningfully intervene in decisions whose input logic cannot be reconstructed (Feretzakis et al., 2025; Sargeant, 2026).

Taken individually, the information provision requirement cannot be satisfied where a model retains no index of its training entries and therefore cannot generate meaningful information about the logic underpinning a specific decision (Feretzakis et al., 2025). The rights to make representations and to contest a decision are similarly hollowed, since an individual who cannot identify which input variables produced an adverse outcome, such as a rejected credit application, has no substantive basis on which to challenge it (Feretzakis et al., 2025; Sargeant, 2026). Human intervention, the safeguard on which the framework relies most heavily, is the most fragile. A reviewer confronting the cascading, unreconstructable decision chains produced when the output of one ADM becomes the input of another can confirm an outcome but cannot genuinely reconstruct its reasoning, reducing intervention to a procedural formality rather than a substantive check (Suksi, 2021). Even the ICO's requirement of regular checks for accuracy and bias is defeated on the same ground, since a system whose reasoning cannot be reconstructed cannot be meaningfully audited for the discrimination those checks are intended to detect (ICO, 2026). The safeguards retained by the DUAA are therefore not merely weakened but structurally mismatched to the technology they are meant to govern, legislating for a transparency the architecture cannot deliver, so that the statutory protections operate in form while failing in substance (Feretzakis et al., 2025; Sargeant, 2026).

3.3 *The Erosion of Privacy in Blurred Sectoral*

Boundaries

While the private sector relies on legitimate interests for data processing, the public sector is required to process personal data on a statutory basis, ensuring that data processing remains subject to democratic oversight and legality (Butler, 2018). Although Article 6 of the GDPR establishes lawful bases for data processing, including consent and legitimate interests, applying these grounds across the public and private sectors introduces both technical limitations and regulatory disparities that weaken individual data protection rights (Butler, 2018; Feretzakis et al., 2025). Once individual data points are transformed into abstract mathematical representations within these models, their origin becomes obscured, and the models lack any index of their training entries, directly undermining individuals' right to know how their data is being processed under Article 15 (Feretzakis et al., 2025).

Scholars have consistently stressed that privacy is context-dependent (Kerr et al., 2009; Kokolakis, 2017; Smith et al., 2011). However, the increasing outsourcing of public functions to private actors has blurred the historic boundary between public authorities and private organisations, creating complex flows of personal data that no longer respect strict sectoral boundaries (Butler, 2018). For instance, DeepMind collaborated with the NHS on the processing of patient records, raising significant privacy concerns (Andreotta et al., 2022). A significant risk emerges from Article 6(1)(f) of the GDPR, whereby private actors performing public functions may elect to rely on the flexible 'legitimate interests' ground rather than the strict statutory legal basis required of public authorities, thereby circumventing the democratic oversight and transparency obligations that should govern the processing of citizens' personal data (Butler, 2018). Scholars and practitioners argue that this regulatory gap is further compounded by the DUAA 2025, which introduces a new lawful

basis of ‘recognised legitimate interests’ under Article 6, removing the balancing test for specified purposes including crime prevention, national security, and responding to emergencies, while codifying direct marketing as an example of an ordinary legitimate interest that remains subject to the standard balancing test (Butler, 2018; Drake & Jameson, 2026).

3.4 Privacy-Preserving Technologies (PETs): Promises and Limitations

Under GDPR Article 25, Privacy by Design and by Default (PbD) mandates that privacy should be proactively built into systems from the design stage and promoted as the default setting (Koops & Leenes, 2014). In response to these regulatory challenges, organisations have increasingly turned to Privacy-enhancing technologies (PETs) (Feretzakis et al., 2025; Miloudi et al., 2026). Several emerging techniques have been developed to help organisations fulfil individuals' right to erasure (Feretzakis et al., 2025; Yan et al., 2025). For instance, Machine Unlearning approaches, including Knowledge Unlearning, enable models to selectively forget sensitive information acquired during initial training, and Offsite Tuning, which transmits only lightweight adapters to data owners to ensure sensitive data is never directly exposed to the full foundation model (Feretzakis et al., 2025; Yan et al., 2025). Beyond unlearning, Homomorphic Encryption enables computation directly on encrypted data, and Secure Multi-Party Computation allows multiple parties to jointly analyse data without exposing their individual inputs, as demonstrated by financial institutions jointly analysing risk models without sharing customer data (Feretzakis et al., 2024).

Federated learning (FL) has also emerged as a promising research area in which models are trained in a decentralised manner, so sensitive data never leaves the local device (Feretzakis et

al., 2024; Miloudi et al., 2026). Unlike deep learning models, which are characterised by centralised learning, where data from multiple sources is aggregated on a single central server for model training (Miloudi et al., 2026). In high-stakes environments such as the health sector, patient data is highly sensitive and subject to strict regulatory protections, making it impractical and often unlawful to pool data into a single centralised system (Miloudi et al., 2026). Federated learning addresses this by allowing AI models to be trained across multiple decentralised data sources, such as hospitals or clinics, without the underlying personal data being transferred to the foundation model central server (Miloudi et al., 2026). A recent success showed that hospitals collaborating to build a predictive model for patient readmissions, where each hospital trains the model locally and only encrypted model updates are sent to a central server, without any data leaving the hospitals (Feretzakis et al., 2024). This makes it a particularly well-suited approach for sectors where privacy demands and data protection obligations cannot be compromised (Miloudi et al., 2026).

Furthermore, Differential Privacy (DP) has emerged as another PET, adding statistical noise during training to protect individual data points (Feretzakis et al., 2025). DP guarantees a limit on the amount of information specific to any single individual that is revealed during an analysis, with noise added either at the data-collection stage or during analysis to a centralised dataset (Feretzakis et al., 2025). The degree of privacy protection is controlled by a tuning parameter called epsilon (ϵ), whereby a lower epsilon value indicates that more noise is added, resulting in stronger privacy protection (Feretzakis et al., 2025).

While these techniques represent meaningful progress toward regulatory compliance, each approach has significant limitations that often introduce new risks (Feretzakis et al., 2025). For

instance, DP incurs considerable performance costs and can inadvertently introduce bias by disproportionately reducing model accuracy for underrepresented data classes, raising serious fairness concerns in high-stakes applications (Feretzakis et al., 2025). Similarly, machine unlearning techniques frequently disrupt model coherence and degrade overall accuracy, creating a persistent tension between legal compliance and technological utility (Feretzakis et al., 2025).

FL, on the other hand, remains vulnerable to data heterogeneity across institutions, where differences in patient demographics and statistical distributions can poison the global model and prevent reliable convergence, directly contributing to algorithmic bias against underrepresented populations in datasets (Miloudi et al., 2026). FL requires high network capacity, which limits the transmission and update of large models in low-resource environments (Miloudi et al., 2026). Compounding these technical limitations is a broader ethical concern; patients rarely know their data is being used to train decentralised models, raising unresolved questions about informed consent that current regulatory frameworks have yet to adequately address (Miloudi et al., 2026).

4. Discussion

The findings of this literature review collectively reveal a mismatch between AI technologies and data protection law (Duffourc et al., 2024; Feretzakis et al., 2025; Janssen, 2020; Sargeant, 2026). Whilst the GDPR was designed to give individuals meaningful control over their personal data, the architectural design of both GenAI and ADMs has rendered several of its core provisions practically unenforceable, exposing a fundamental misalignment between regulatory intent and technical reality (Feretzakis et al., 2025; Sargeant, 2026). Across this literature, scholars converge on the view

that neither technical fixes nor new AI-specific statutes suffice on their own; closing the gap instead requires a layered, lifecycle-oriented programme of reform spanning the legal, institutional, and design dimensions of how these systems are governed (Feretzakis et al., 2024; Miloudi et al., 2026).

Underlying each of these dimensions is a structural asymmetry of knowledge and power between corporations and the individuals whose data they exploit, which remains a foundational barrier to effective privacy protection (Kokolakis, 2017). Informed consent is not merely a procedural transfer of information but carries symbolic value that acknowledges a person's agency and rights (Andreotta et al., 2022). Yet three persistent problems continue to hollow out meaningful consent in the AI age; a lack of transparency about how data is used, the repurposing of data beyond its original scope, and the absence of meaningful alternatives for users wishing to protect their privacy (Andreotta et al., 2022). Access to services frequently forces users to accept all terms and conditions or lose access entirely, exposing them to wider privacy risks (Barth et al., 2022; Whitley & Pujadas, 2018), and empirical evidence shows that GenAI browser extensions routinely track and share user queries and chat histories with third-party trackers in direct contradiction of their own privacy policies (Vekaria et al., 2025). Because users rely on the broader regulatory environment rather than reading terms and conditions, the obligation to ensure transparent and fair data practices rests with firms rather than individuals (Whitley & Pujadas, 2018).

The first dimension concerns the reintegration of enforceable legal accountability into the design and deployment of ADMs. The UK's soft-governance approach has produced a vicious cycle in which the discovery of non-compliant systems erodes public trust, which in turn reduces organisational motivation for transparency and fuels further legal contestation

and public backlash (Drake et al., 2022). To break this cycle, researchers call for structural policy reforms that mandate the publication of Data Protection Impact Assessments and preserve a meaningful right to explanation (Drake et al., 2022). In high-stakes contexts, interpretable models such as decision trees and rule-based systems should be the default choice in ADMs, given their inherent explainability and capacity to support individuals' rights to understand and contest decisions affecting their access to services (Sargeant, 2026). Although their adoption in the public sector disrupts established accountability frameworks, and therefore requires clear legal mechanisms for assigning liability when errors occur (Suksi, 2021). Holtz and Ledendal (2026) operationalise these principles into a four-benchmark fundamental rights impact assessment embedded within the GDPR's DPIA process, with mitigation escalating from human-in-the-loop review and continuous audits for lower-risk systems to sunset clauses or moratoria for high-risk deployments.

The second dimension concerns institutional and regulatory architecture, where scholars recommend that the UK expand and fund existing data protection authorities such as the ICO, rely on the UK GDPR's existing automated decision-making provisions, and use the Digital Regulation Cooperation Forum to coordinate sectoral oversight (McDougall, 2023), complemented by AI-lifecycle-specific regulatory guidance and legally mandated AI ethics officers to oversee transparency and mitigate algorithmic bias (Mbonye et al., 2024). At the organisational level, Feretzakis et al. (2025) propose a four-layer, GDPR-aligned governance framework that combines lawful-basis and data-minimisation controls with privacy-enhancing techniques such as differential privacy, federated learning, and modular architectures for targeted unlearning, stressing that alignment ultimately depends on a "GDPR-first" organisational culture rather than

technology alone.

The third dimension concerns design, and in particular the reframing of Privacy by Design. Although PbD is intended to embed security, transparency, and accountability into systems from the outset (Barth et al., 2022), its effectiveness has been undermined by the widespread misconception that it simply requires developers to encode GDPR provisions directly into software (Koops & Leenes, 2014). In practice, developers must simultaneously navigate overlapping jurisdictions and domain-specific legislation, a task for which uniform technical encoding is impossible; regulators must therefore reframe PbD not as a procedural compliance tool imposed on engineers but as a substantive, communicative practice that educates and shifts the internal culture of the institutions commissioning these systems (Koops & Leenes, 2014). Complementing this cultural shift, the UK Government's Algorithmic Transparency Recording Standard provides the transparency infrastructure to render such reforms auditable, requiring organisations to document both technical specifications and non-technical governance risks for public and expert scrutiny (Kingsman et al., 2022). By piloting the Standard in the public sector first, the government aims to establish an ecosystem of trust that drives industry-wide adoption (Kingsman et al., 2022); its transformative potential, however, remains contingent on voluntary private-sector engagement, particularly as many companies assert compliance with their legal obligations without adequately disclosing the technical safeguards underpinning those claims (Feretzakis et al., 2025; Kingsman et al., 2022).

5. Conclusion

The architectural design of modern deep learning systems has created severe practical barriers to the exercise of core data subject rights, including the rights to erasure, rectification, and

explanation, exposing a fundamental and widening socio-technical misalignment between the regulatory framework and the technical reality it seeks to govern (Duffourc et al., 2024; Feretzakis et al., 2025; Sargeant, 2026). Scholars argue that the post-Brexit implementation of the DUAA has deepened this crisis by permitting solely automated decision-making in wider circumstances and by expanding the legitimate interests basis for data processing, prioritising economic growth over individual rights at a moment when stronger safeguards are most urgently needed, although the government maintains that the Act retains the UK GDPR and preserves statutory safeguards (GOV.UK, 2025; Sargeant, 2026). Privacy-preserving technologies offer partial pathways to compliance; however, each has significant limitations and raises unresolved questions about informed consent, confirming that technical solutions alone cannot bridge the regulatory gap (Feretzakis et al., 2025; Miloudi et al., 2026; Yan et al., 2025).

The AI field is still growing, and more research is needed to address these privacy concerns (Feretzakis et al., 2025; Miloudi et al., 2026). Achieving absolute privacy today has proven mathematically impossible, leaving organisations with a greater legal responsibility to maintain transparency and deploy reasonable safeguards rather than guarantee perfect protection (Feretzakis et al., 2024, 2025; Kingsman et al., 2022). The legal dimension requires reintegrating enforceable accountability into ADM deployment through the mandatory publication of Data Protection Impact Assessments and the preservation of a meaningful right to explanation (Drake et al., 2022). A default preference for interpretable models in high-stakes automated decisions so that the DUAA's statutory safeguards can operate as intended (Sargeant, 2026). The institutional dimension calls for strengthening and funding existing regulators such as the ICO and coordinating them through the Digital

Regulation Cooperation Forum (McDougall, 2023), supported by AI-lifecycle-specific guidance, legally mandated ethics officers, and a "GDPR-first" organisational culture rather than reliance on technology alone (Feretzakis et al., 2025; Mbonye et al., 2024). The design dimension requires reframing Privacy by Design as a cultural and communicative practice rather than a compliance exercise (Koops & Leenes, 2014), underpinned by sector-wide adoption of the Algorithmic Transparency Recording Standard as the transparency infrastructure that renders the other reforms auditable (Kingsman et al., 2022). The Standard is necessary but not sufficient; as scholars noted, its transformative potential remains contingent on voluntary private-sector engagement, and transparency without enforceability would reproduce the very gap this review has documented (Kingsman et al., 2022).

References

- Akakpo, A., Gyasi, E. A., Oduro, B., & Akpabot, S. (2025). Artificial Intelligence (AI) and the Future of Information Privacy: Expert Viewpoints. *Journal of Global Information Management*, 33(1), 1–63:1-63:37.
25. <https://doi.org/10.4018/JGIM.383050>
- Aldemir, C., & Uysal, T. U. (2025). Artificial Intelligence for Financial Accountability and Governance in the Public Sector: Strategic Opportunities and Challenges. *Administrative Sciences*, 15(2), 58.
<https://doi.org/10.3390/admsci15020058>
- Andreotta, A. J., Kirkham, N., & Rizzi, M. (2022). AI, big data, and the future of consent. *AI & SOCIETY*, 37(4), 1715–1728.
<https://doi.org/10.1007/s00146-021-01262-5>
- Anic, I.-D., Budak, J., Rajh, E., Recher, V., Skare, V., & Skrinjaric, B. (2019). Extended model of online privacy concern: What drives consumers' decisions? *Online Information Review*, 43(5), 799–817.
<https://doi.org/10.1108/OIR-10-2017-0281>
- Barth, S., Ionita, D., & Hartel, P. (2022). Understanding Online Privacy—A Systematic Review of Privacy Visualizations and Privacy by Design Guidelines. *ACM Comput. Surv.*, 55(3), <https://doi.org/10.1145/3502288>
- Bonina, C., Koskinen, K., Eaton, B., & Gawer, A. (2021). Digital platforms for development: Foundations and research agenda. *Information Systems Journal*, 31(6), 869–902.
<https://doi.org/10.1111/isj.12326>
- Butler, O. (2018). Obligations Imposed on Private Parties by the GDPR and UK Data Protection Law: Blurring the Public-Private Divide. *European Public Law*, 24(3).
<https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\EURO\EURO2018031.pdf>
- D'Hauwers, R. (2026). *Data brokers market study*. European Data Protection Board.
https://www.edpb.europa.eu/our-work-tools/our-documents/support-pool-experts-projects/data-brokers-market-study_en
- Drake, A., Keller, P., Pietropaoli, I., Puri, A.,

- Maniatis, S., Tomlinson, J., Maxwell, J., Fussey, P., Pagliari, C., Smethurst, H., Edwards, L., & Blair, S. W. (2022). Legal contestation of artificial intelligence-related decision-making in the United Kingdom: Reflections for policy. *International Review of Law, Computers & Technology*, 36(2), 251–285.
<https://doi.org/10.1080/13600869.2021.1999075>
- Drake, E., & Jameson, A. (2026, February 3). *UK GDPR UK privacy reform is finally going live what does your business need to do now—Bird & Bird*. Bird & Bird.
<https://www.twobirds.com/en/insights/2026/uk/uk-gdpr-uk-privacy-reform-is-finally-going-live--what-does-your-business-need-to-do-now>
- Duffourc, M. N., Gerke, S., & Kollnig, K. (2024). *Privacy of Personal Data in the Generative AI Data Lifecycle – NYU Journal of Intellectual Property & Entertainment Law*.
<https://jipel.law.nyu.edu/privacy-of-personal-data-in-the-generative-ai-data-lifecycle/>
- Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., & Verykios, V. S. (2024). Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review. *Information*, 15(11), 697.
<https://doi.org/10.3390/info15110697>
- Feretzakis, G., Vagena, E., Kalodanis, K., Peristera, P., Kalles, D., & Anastasiou, A. (2025). *GDPR and Large Language Models: Technical and Legal Obstacles*. *Future Internet*, 17(4), 151.
<https://doi.org/10.3390/fi17040151>
- GOV.UK. (2025). Data (Use and Access) Act 2025: Data protection and privacy changes. Department for Science, Innovation and Technology.
<https://www.gov.uk/guidance/data-use-and-access-act-2025-data-protection-and-privacy-changes>
- Holtz, H. M., & Ledendal, J. (2026). AI data governance – overlaps between the AI Act and the GDPR. *Law, Innovation and Technology*, 18(1), 380–409.
<https://doi.org/10.1080/17579961.2026.2633677>
- Information Commissioner's Office (ICO). (2026). The Data (Use and Access) Act 2025 (DUAA): What does it mean for organisations?
<https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/data-use-and-access-act-2025/>

- Janssen, H. L. (2020). An approach for a fundamental rights impact assessment to automated decision-making. *International Data Privacy Law*, 10(1), 76–106. <https://doi.org/10.1093/idpl/ipz028>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kerr, I., Lucock, C., & Steeves, V. (2009). *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society*. Oxford University Press, Incorporated. <http://ebookcentral.proquest.com/lib/londonschool/olecons/detail.action?docID=3054189>
- Kingsman, N., Kazim, E., Chaudhry, A., Hilliard, A., Koshiyama, A., Polle, R., Pavey, G., & Mohammed, U. (2022). Public sector AI transparency standard: UK Government seeks to lead by example. *Discover Artificial Intelligence*, 2(1), 2. <https://doi.org/10.1007/s44163-022-00018-4>
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122–134. <https://doi.org/10.1016/j.cose.2015.07.002>
- Koops, B.-J., & Leenes, R. (2014). Privacy regulation cannot be hardcoded. A critical comment on the ‘privacy by design’ provision in data-protection law. *International Review of Law, Computers & Technology*, 28(2), 159–171. (95888802). <https://doi.org/10.1080/13600869.2013.801589>
- Kshetri, N. (2025). Governing Agentic AI: Security, Identity, and Oversight in the Age of Autonomous Intelligent Systems. *Computer*, 58(8), 123–129. <https://doi.org/10.1109/MC.2025.3572173>
- Lermen, S., Paleka, D., Swanson, J., Aerni, M., Carlini, N., & Tramèr, F. (2026). *Large-scale online deanonymization with LLMs* (arXiv:2602.16800). arXiv. <https://doi.org/10.48550/arXiv.2602.16800>
- Lim, S., & Shim, H. (2022). No secrets between agents. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 16(4).

<https://doi.org/10.5817/CP2022-4-3>

MacArthur, B. D., Dorobantu, C. L., & Margetts, H. Z. (2022). Resilient government requires data science reform. *Nature Human Behaviour*, 6(8), 1035–1037. <https://doi.org/10.1038/s41562-022-01423-6>

Majeed, A., & Hwang, S. O. (2023). When AI Meets Information Privacy: The Adversarial Role of AI in Data Sharing Scenario. *IEEE Access*, 11, 76177–76195. <https://doi.org/10.1109/access.2023.3297646>

Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *J. Strateg. Inf. Syst.*, 30(3). <https://doi.org/10.1016/j.jsis.2021.101683>

Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature*, 568(7751), 163–165. <https://doi.org/10.1038/d41586-019-01099-5>

the two of us: Privacy concerns over using AI

McDougall, S. (2023). More Speed, Less Haste: Finding an Approach to AI Regulation that Works for the UK. *Amicus Curiae*, 5(1), 104–125. <https://doi.org/10.14296/ac.v5i1.5664>

Meng, X., & Liu, J. (2025). “Talk to me, I’m secure”: Investigating information disclosure to AI chatbots in the context of privacy calculus. *Online Information Review*, 49(5), 933–954. <https://doi.org/10.1108/OIR-06-2024-0375>

Miloudi, A., Laouid, A., Bounceur, A., Kara, M., Bouhamed, M. M., Hammoudeh, M., & Kraidia, I. (2026). Federated learning in healthcare: Recent progress and challenges. *Computers and Electrical Engineering*, 131, 110924. <https://doi.org/10.1016/j.compeleceng.2025.110924>

Murray, A. (2010). *Information Technology Law: The Law and Society*. OUP Oxford.

Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57(6), 1701–1778.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews.

Information & Management, 52(2), 183–199.

<https://doi.org/10.1016/j.im.2014.08.008>

Sargeant, H. (2026). Mind the gap: Securing algorithmic explainability for credit decisions beyond the UK GDPR. *Computer Law & Security Review*, 60, 106247.

<https://doi.org/10.1016/j.clsr.2025.106247>

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, 43(3), 695–719.

<https://doi.org/10.25300/MISQ/2019/13747>

Smith, H. J., Dinev, T., & Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly*, 35(4), 989–1015. <https://doi.org/10.2307/41409970>

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3), 477–564. <https://doi.org/10.2307/40041279>

Suksi, M. (2021). Administrative due process when using automated decision-making in public administration: Some notes from a Finnish perspective. *Artificial Intelligence and Law*,

29(1), 87–110.

<https://doi.org/10.1007/s10506-020-09269-x>

UK GDPR. (2021). *UK GDPR*. UK GDPR | Fieldfisher. <https://ukgdp.fieldfisher.com/>

UKAN. (2019, April 26). Anonymisation with differential privacy [UK Anonymisation Network]. *Anonymisation with Differential Privacy*. <https://ukanon.net/framework/>

Vekaria, Y., Canino, A. L., Levitsky, J., Ciechonski, A., Callejo, P., Mandalari, A. M., & Shafiq, Z. (2025). Big help or big brother? Auditing tracking, profiling, and personalization in generative AI assistants. Proceedings of the 34th USENIX Security Symposium, 8115–8134.

<https://doi.org/10.5555/3766078.3766494>

Vogelsang, A., & Borg, M. (2019). Requirements Engineering for Machine Learning: Perspectives from Data Scientists. *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, 245–251.

<https://doi.org/10.1109/REW.2019.00050>

Wang, J., Kiran, E., Aurora, S. R., Simeone, M., & Lobo, J. (2025). ChatGPT on ChatGPT: An Exploratory Analysis of its Performance in the

Public Sector Workplace. *Digit. Gov.: Res. Pract.*, 6(2), 29:1-29:28.

<https://doi.org/10.1145/3676281>

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii–xxiii.

Whitley, E. A., & Pujadas, R. (2018). Report on a study of how consumers currently consent to share their financial data with a third party. In *Financial Conduct Authority (FCA): London, UK*. [Report]. Financial Conduct Authority (FCA).

<https://www.fca.org.uk/panels/consumer-panel/publication/report-study-how-consumers-currently-consent-share-their-financial-data-third-party>

Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., data privacy data privacy of Large Language Models (LLMs) and LLM agents: A literature review. *High-Confidence Computing*, 5(2), 100300. <https://doi.org/10.1016/j.hcc.2025.10>

