cerre

MAKING DATA PROTECTION FIT FOR THE AGE OF AI

REPORT

June 2026

Marco Bassini
Cristiana Firullo

As provided for in CERRE's bylaws and procedural rules from its "Transparency & Independence Policy", all CERRE research projects and reports are completed in accordance with the strictest academic independence.

The project, within the framework of which this report has been prepared, received the support and/or input of the following CERRE member organisations: Amazon, Apple, CnaM, Microsoft, Mozilla. However, they bear no responsibility for the contents of this report. The views expressed in this CERRE report are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, they do not necessarily correspond either to those of CERRE, or of any sponsor or of members of CERRE.

info@cerre.eu – www.cerre.eu

Executive Summary

This report examines the Digital Omnibus, the legislative package proposed by the European Commission in November 2025 (COM/2025/837 and COM/2025/836), as both a legal and an economic intervention in the EU digital rulebook. Its proposed simplification measures recalibrate the framework governing the interaction between data protection and artificial intelligence. We ask whether the measures do so in a manner that is both innovation-oriented and consistent with the fundamental rights commitments of EU law. The report combines doctrinal legal analysis with a law-and-economics perspective and, where relevant, a law-and-technology lens, and it concludes with recommendations addressed to EU lawmakers and organised around a three-phase reform roadmap. We conclude that the Omnibus is a constructive attempt to make EU data protection fit for the age of AI. However, to help Europe make the most of AI, lawmakers must clarify the law and ensure changes do not hurt smaller businesses, thus stifling competition.

The report is framed by a deliberately non-adversarial hypothesis: data protection and AI-driven innovation are not inherently in conflict. In a substantial range of settings, such as personalisation, bias mitigation, linguistic and cultural diversity, automated compliance, and privacy-enhancing technologies, the interests of data subjects and AI developers are structurally aligned. The challenge is not to choose between privacy and innovation, but to design a framework capable of distinguishing scenarios in which these interests converge from those in which they genuinely diverge. It is against that standard that we assess each provision of the Digital Omnibus, without any general assumptions about the optimal amount of data protection. Three genuine gains stand out from our analysis:

- The codification of legitimate interest as a legal basis for AI model training (Article 88c GDPR) removes significant legal uncertainty that has hindered European AI development. The Omnibus confirms that consent is not the only basis for large-scale training. However, various parts of the Omnibus risk reintroducing fragmentation and uncertainty[1]
- The Omnibus allows the use of data to detect and mitigate bias in all AI systems, not just high-risk ones, and not just for providers of the AI system (Article 4a of the Digital Omnibus on AI). This is a sound response to evidence that algorithmic bias is not confined to high-risk applications, while its strict-necessity requirement keeps the processing of special categories of data properly circumscribed.
- The shift from the ePrivacy Directive to the GDPR of rules for the storing of, and access to, data on terminal equipment (i.e., cookies) (Article 88a GDPR) is a promising move towards a more flexible, risk-based, and future-proof regime, though it requires broader exemptions for genuinely low-risk processing and a reconsideration of rigid features such as the proposed six-month ban on renewed consent requests.

Alongside these gains, the report identifies a recurring pattern of provisions whose effects cut against their stated objectives:

[1] These include ambiguous wording (notably “where appropriate”), the undefined notion of the “operation” of AI, the omission of third-party interests, and the scope for Member States to reintroduce a consent requirement.

- The proposed redefinition of personal data (Article 4(1) GDPR), codifying a relative, entity-specific approach, was rejected following the joint EDPB-EDPS opinion. This confirms that **efforts to facilitate innovation should not come through amendments to the GDPR's foundational concepts. Such changes would directly affect the scope of protection of the fundamental right to data protection in the EU legal system and would require careful assessment of their broader impact.** Instead, this requires a more refined recalibration of existing mechanisms.
- The enforcement of data subject rights — particularly the right to erasure — in trained AI models raises a **structural mismatch between a right premised on identifiable, reversible operations and the distributed nature of machine learning**. Compliance cannot be structured as an ex-post remedy but must be built into model design, a conclusion that calls for guidance grounded in technical feasibility rather than additional administrative cost.
- **The restriction on the right to access one's personal data (Article 12(5) GDPR)** has attracted strong criticism; the report finds these concerns partly overstated but recommends refining the amendment to preserve the contestation-enabling function of the right of access in AI-driven contexts.
- **The centralised consent-management and privacy-signals mechanism (Article 88b GDPR) is a proposal with ambiguous theoretical effects and potentially disruptive consequences for the online ecosystem**. While it might reduce the number of consent prompts, it could spell the end of a vast amount of ad-supported free online content, and lead to more paywalls for European readers. Moreover, there is no evidence that it would yield better-informed signals, and it raises serious questions of legal validity under Article 4(11) GDPR, standardisation feasibility, and competition, given the concentration of the browser market.

A consistent economic finding runs across these provisions: the GDPR's largely fixed compliance costs fall disproportionately on small and medium-sized enterprises and independent AI developers. Several Omnibus measures, despite their stated SME-supportive objectives, risk consolidating the position of large incumbents — those best placed to absorb iterative compliance assessments and to internalise proprietary data infrastructures — thereby partially offsetting the simplification gains the package articulates. On this basis, the report recommends a three-phase approach:

- **In the context of the Digital Omnibus**, priority should be given to clarifying and refining the proposed amendments to reduce legal uncertainty without altering the core structure of the GDPR.[2]
- **Within the Digital Fitness Check**, the European Commission needs to make a more systematic effort to ensure coherence across the EU digital acquis, particularly the interaction between the GDPR and the AI Act on risk management, data governance, and the allocation of responsibilities across the AI lifecycle.

[2] Specifically, this would include specifying Article 88c GDPR, framing the Article 9(5) GDPR "avoid–remove–protect" logic as a lifecycle obligation, and ensuring that data-subject-rights obligations remain technically realistic.

- **In the longer term**, a targeted, risk-based adaptation of the GDPR — supported by harmonised guidance rather than wholesale reform — should better distinguish processing that serves the interests of both users and developers from processing that creates genuine risks to fundamental rights.

In conclusion, the Digital Omnibus represents a timely and, in important respects, constructive attempt to make EU data protection fit for the age of AI. Yet, its benefits will materialise only if the negotiation phase resolves some interpretive ambiguities and attends to the distributional incidence of the reforms on smaller market participants, beyond treating structural measures, such as the privacy-signals mechanism, as interventions to be carefully assessed before adoption.

Table of Contents

1. Introduction

The European Union's data protection framework[3] occupies a central place in the global regulatory landscape. Its rules have shaped the development of the EU digital economy for more than a decade and have exercised a recognisable extraterritorial influence — what the literature has termed the *Brussels effect*[4] – on jurisdictions far beyond the Union's borders. Yet the same framework has come under sustained criticism for imposing regulatory complexity on firms operating in the EU – including a fragmented implementation across Member States, a disproportionate compliance burden placed on small and medium-sized enterprises, and friction between the protection of fundamental rights and the imperatives of innovation in data-intensive technologies. This friction is most notable in the development of AI systems whose performance depends on access to large, diverse, and representative datasets. These negative aspects have emerged as distinguishing traits of the European Union paradigm in the governance of digital technologies, which prioritises regulation but increasingly lacks technological leadership, which we see e.g., in China and the United States[5]. These concerns animated the recommendations of two recent reports – *The Future of European Competitiveness*[6] and *Much More Than a Market*[7] – and informed the European Commission's decision to introduce, in November 2025, a legislative package commonly referred to as the Digital Omnibus. The package, comprising two proposals for regulations – COM/2025/837, the "Digital Omnibus", and COM/2025/836, also known as "Digital Omnibus on AI" – are currently under negotiation between the European Parliament and the Council. The proposals would amend selected provisions of the GDPR and other digital laws.[8] The package's stated objective is to reduce administrative burdens, lower compliance costs – particularly for SMEs – and create a regulatory environment conducive to innovation, while preserving the EU's high standards of fundamental rights protection.

This paper examines the Digital Omnibus as both a legal and an economic intervention. Its central question is whether the proposed simplification measures recalibrate the EU digital rulebook in a manner that is *both* innovation-oriented and consistent with the fundamental rights commitments of EU primary law. The analysis is framed by the core hypothesis that data protection and AI-driven innovation are not inherently in conflict. As noted by the European Data Protection Board (EDPB) in O[9], "AI technologies create many opportunities and benefits across a wide range of sectors and social activities" and the "GDPR is a legal framework that encourages responsible innovation". In a substantial range of settings, the interests of data subjects and AI developers are structurally aligned – users stand to benefit directly from the development of more capable AI systems, they increasingly demand personalised AI-driven services, and the need for diverse, representative datasets serves both the mitigation of algorithmic bias and the improvement of model performance for populations whose representation in training data has historically been incomplete. The challenge lies not in choosing between privacy and innovation, but in the inability of the current regulatory framework to

[3] Embodied in the General Data Protection Regulation (Regulation (EU) 2016/679, hereinafter "GDPR") and other pieces of legislation (including the more recent Artificial Intelligence Act, i.e., Regulation (EU) 2024/1689, hereinafter "AI Act").
[4] Bradford A. (2020). The Brussels Effect: How the European Union Rules the World. Oxford University Press.
[5] Bradford A. (2023). Digital Empires. The Global Battle to Regulate Technology. Oxford University Press.
[6] Draghi, M. (2024). The future of European competitiveness.
[7] Letta, E. (2024). Much More Than a Market. Speed, Security, Solidarity. Empowering the Single Market to deliver a sustainable future and prosperity for all EU Citizens.
[8] These include the Data Act (Regulation (EU) 2023/2854), the Data Governance Act (Regulation (EU) 2022/868), the NIS2 Directive (Directive (EU) 2022/2555), the ePrivacy Directive (Directive 2002/58/EC), the Critical Entities Resilience (CER) Directive (Directive (EU) 2022/2557), and the AI Act (Regulation (EU) 2024/1689).
[9] EDPB (2024a). Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models.

adequately differentiate between scenarios in which these interests converge and those in which they genuinely diverge. In this way, the current framework generates unnecessary friction in both directions and produces outcomes that are simultaneously over-inclusive, in that they burden processing activities that serve user interests, and under-protective, in that they rely on instruments whose cognitive and informational preconditions are systematically unmet.

Against this background, the paper pursues three research questions.

First, does the Digital Omnibus deliver on its stated simplification objectives, and at what cost in terms of legal certainty and fundamental rights protection? This question is not exclusively a matter of business choices for the private sector: thinking of data protection authorities that more and more will have to enforce the GDPR in AI-mediated data processing activities, regulators may face legal uncertainty and thus incur additional enforcement costs. Second, do the proposed amendments adequately distinguish between scenarios of interest convergence and divergence between data subjects and AI developers, or do they replicate, in altered form, the undifferentiated approach of the existing framework? Third, what refinements to the proposals — both within the current legislative cycle and in the broader reform agenda announced by the European Commission — would best support a regulatory environment that is innovation-oriented, rights-protective, and proportionate in its incidence across the European digital ecosystem, including small and medium-sized firms and independent AI developers?

The methodology adopted to address these questions integrates doctrinal legal analysis with law-and-economics perspectives and, where substantive issues require it, draws on a law-and-technology perspective to examine tensions between legal requirements and technical feasibility – for example, in assessing the practical enforceability of the right to erasure in trained AI models. First, *doctrinal legal analysis* – the systematic analysis of the proposed amendments in light of the existing GDPR and AI Act, the case law of the Court of Justice, the relevant decisions of supervisory authorities and the guidance of the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) – provides the framework within which we assess the legal coherence of the Digital Omnibus package. Second, *law-and-economics analysis* draws on empirical research in the economics of privacy, AI, and digital markets to evaluate the distributional incidence and the welfare implications of the proposed amendments, with particular attention to their effects on SME competitiveness, innovation incentives, and the European AI ecosystem. Third, where substantive issues require it, we adopt a *law-and-technology* perspective to examine tensions between legal requirements and technical feasibility – for example, in assessing the practical enforceability of the right to erasure in trained AI models. The analysis only covers the provisions of the Digital Omnibus most relevant to the interaction between data protection and AI; the limits and normative assumptions of this scope are made explicit in each substantive subsection. This way, the report aims to address the most significant amendments proposed as part of the Digital Omnibus and their expected impact on both innovation and fundamental rights protection, while it also proposes longer-term recommendations relating to AI developments and data protection rules. In other words, the discussion on the amendments proposed in the Digital Omnibus provides a timely opportunity to reflect upon highly disputed legal issues. The paper will therefore also determine the extent to which the Digital Omnibus is likely to deliver (or fail to deliver) on its promises.

The structure of the paper reflects the complementary character of the analysis: Part I situates the Digital Omnibus within the policy and regulatory context that produced it (Sections 2 and 3); Part II provides the substantive legal and economic analysis of the proposed amendments (Section 4); and Part III offers an aggregate assessment and a set of recommendations addressed to EU lawmakers, organised around a three-phase reform roadmap (Section 5).

Part I

2. Moving ahead: the quest for change

The relationship between data protection and AI-driven innovation is frequently framed as a trade-off: stronger privacy safeguards are assumed to constrain the data access that AI systems require, while weaker protections are thought to enable faster technological progress. Yet, these framing obscures a more nuanced reality. In a wide range of settings, the interests of users and AI developers are not opposed but structurally aligned. Users benefit from more capable AI models — models that deliver personalised recommendations, more accurate translations, and context-sensitive services — and developers, in turn, require large-scale, diverse datasets to build systems that perform well and serve heterogeneous populations equitably. The challenge, then, is not simply to balance privacy against innovation, but to identify when and where these interests converge, when they genuinely diverge, and whether the regulatory framework is equipped to distinguish between the two.

This section examines that challenge through the lens of empirical research in economics, computer science, and the behavioural sciences. It develops a two-layered analysis that mirrors the dual structure of the problem. From the users' perspective, the evidence reveals a set of compounding obstacles – cognitive overload, systematic behavioural biases, persistent informational opacity – that undermine the effectiveness of the consent-based model at the heart of EU data protection law, even as users increasingly demand the personalised services that data processing enables. From the firms' perspective, a growing body of empirical work documents the disproportionate compliance costs borne by smaller firms, the dampening effects on venture investment and market entry, and the critical role that data diversity plays in ensuring model quality, mitigating algorithmic bias, and representing the EU's cultural and linguistic plurality.

2.1. Users' perspective

In the data economy, personal data can function as a form of currency. Users can access digital services without necessarily paying monetary prices (e.g., search engines, social media platforms, AI-powered assistants, and recommendation systems). The transaction takes a different form: users exchange access to their personal information for access to services, and firms use that information to fund their operations through targeted advertising, to improve their products through behavioural analytics and model training, and to deliver the personalised experiences that users have come to expect. To a large part of the data economy, this exchange is not incidental; it is constitutive of it[10].

Most of the AI-driven services that characterise the current digital landscape – from real-time translation to medical symptom checkers to personalised learning platforms – depend on the availability of large volumes of personal data both for their initial development and for their ongoing operation. Understood as an exchange, the circulation of personal data generates substantial benefits on the user side. More data enables more capable AI models, which in turn deliver services of higher quality and greater relevance. Personalisation – the ability of a service to adapt to an individual user's

[10] It is worth noticing that in other parts of the data economy, however, personal data is not used as currency – think for example of B2B cloud services where data is not commoditised but rather highly protected. In other words, different models and variations co-exist.

preferences, context, and history – is not a mere commercial convenience but a feature that users actively demand and that, in many domains, significantly improves outcomes. A recommendation algorithm trained on richer data produces better suggestions; a language model trained on diverse datasets performs more accurately across languages and cultural contexts; a diagnostic tool trained on broader clinical data generates more reliable assessments. In these settings, the user's decision to contribute data is not a sacrifice of privacy in exchange for nothing, but rather a transaction in which the user receives tangible and often substantial value. And yet, when users are asked directly about their preferences regarding data sharing, a strikingly different picture emerges. The effectiveness of this exchange depends on a condition that, as we document below, is systematically unmet: that users can evaluate what they are giving up. This subsection reviews the evidence on the cognitive biases, informational asymmetries, and conceptual confusions that prevent users from accurately assessing the terms of the transaction, and that call into question the reliability of both revealed and stated privacy preferences as a guide for regulatory design.

The first well-documented phenomenon in the privacy literature is the so-called "privacy paradox". Survey evidence consistently shows that large majorities of users express concern about how their personal data is collected and used, report a desire for greater control over their information, and state that they would prefer more restrictive data practices[11]. The gap between these stated preferences and actual behaviour – what the literature has termed the "privacy paradox" – is one of the most extensively documented phenomena in the empirical study of privacy. Users who declare strong privacy concerns continue to use data-intensive services, accept default consent settings, and share personal information in contexts where they could, in principle, decline to do so. The conventional framing presents two broad interpretations of this paradox. The first is that users are *strategically inattentive*, namely they understand the nature of the exchange but prefer not to dwell on its implications, expressing the socially desirable preference for more privacy when asked while implicitly accepting the terms of the transaction through their behaviour. The second is that users genuinely *do not understand the exchange* – that they lack the information, the cognitive resources, or both to assess what they are giving up, what they are receiving, and whether the terms are fair. Before adjudicating between these interpretations, however, it is worth noting a more fundamental problem with the evidence on which they both rely, namely that users' stated privacy preferences are considerably less coherent than they appear. Recent empirical work has shown that the survey scales used to elicit privacy preferences fail to reliably distinguish between conceptually distinct constructs (such as privacy attitude, preference, concern, expectation, decision, and behaviour) even when participants are asked to do so directly[12]. The implication is that aggregate measures of "privacy concern", on which the privacy-paradox argument largely rests, may be capturing a heterogeneous mix of constructs rather than a stable underlying preference. Complementary evidence documents the existence of a reverse privacy paradox: individuals who declare themselves dismissive of privacy nonetheless engage in behaviours that meet established definitions of privacy-protective conduct[13]. Both findings converge on the diagnosis we develop here: "privacy" is not a single object of preference but an umbrella term aggregating distinct demands whose terms of trade vary across users and

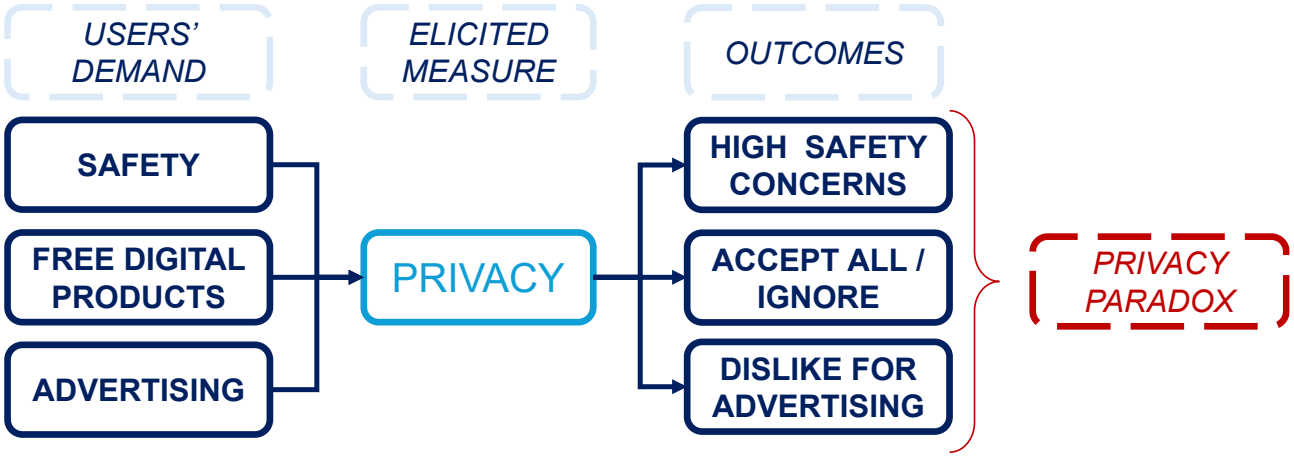
[11] Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and Human Behavior in the Age of Information. Science, 347(6221), 509–514.

[12] Colnago, J., Cranor, L. F., Acquisti, A., & Stanton, K. H. (2022). Is it a concern or a preference? An investigation into the ability of privacy scales to capture and distinguish granular privacy constructs. Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022), 331–346.

[13] Colnago, J., Cranor, L. F., & Acquisti, A. (2023). Is there a reverse privacy paradox? An exploratory analysis of gaps between privacy perspectives and privacy-seeking behaviors. Proceedings on Privacy Enhancing Technologies, 2023(1), 455–476.

contexts. Figure 1 illustrates this conflation: three analytically distinct user demands – safety, demand for free digital products, and attitudes toward advertising – are aggregated under the single elicited measure of 'privacy', producing observed outcomes that the literature interprets as a paradox.

Figure 1 The privacy paradox as an improper aggregation of different effects to the same set of causes



What surveys and consent interactions capture under the label of "privacy concern" typically conflates three analytically distinct demands that are routinely bundled under the single heading. The first is a preference regarding marketing and personalisation – whether and to what extent users want their data to be used to tailor commercial communications, recommendations, and service experiences. The second is a concern about data safety – whether personal information is adequately protected against unauthorised access, breaches, and misuse by third parties. The third is a demand for free or subsidised digital products, such as search engines and social media platforms. From the user's perspective, these are fundamentally different problems. A user who objects to receiving targeted advertisements is expressing a preference about a business model; a user who fears that a data breach will expose sensitive health or financial information is expressing a concern about security. Yet in surveys, consent interfaces, and public discourse, the two are routinely bundled under the single heading of "privacy", and it is rarely clear which dimension drives any given expression of concern – or whether the user even distinguishes between them. This conflation is understandable, because marketing preferences and data safety, while distinct from the user's standpoint, share common roots at the level of firm implementation. Both depend on the same underlying infrastructure of data collection, storage, access controls, and cybersecurity practices. A firm's decisions about how it manages personal data simultaneously determine the degree of personalisation it can offer and the level of security risk its users face. From a regulatory standpoint, however, the appropriate interventions differ substantially. Marketing-related preferences call for mechanisms of user choice and transparency – consent management, opt-out tools, machine-readable privacy signals – instruments that presuppose an informed and engaged user. Data safety concerns, by contrast, call for obligations on firms regarding security standards, breach notification, and accountability – interventions whose effectiveness does not depend on individual user decisions at all. By treating privacy as a unitary concept, the current framework risks applying choice-based instruments to problems that require security-based solutions, and vice versa. It also means that aggregate expressions of privacy concern – which may primarily reflect anxiety about breaches and security

failures – are interpreted as mandates for more restrictive consent requirements, which address a different problem entirely. This measurement problem compounds the interpretive difficulties surrounding the privacy paradox, but the weight of the evidence nonetheless points toward a clear conclusion: whatever the mix of strategic inattention and genuine incomprehension, current privacy choices do not reflect well-informed preferences. Several mutually reinforcing factors account for this.

2.1.1 Consent fatigue

The most immediate obstacle to informed privacy choice is the sheer volume of decisions the current framework demands. The European data protection framework aims at placing the individual in the position of evaluating privacy trade-offs at the point of data collection. While the GDPR itself relies on six legal bases set out in Article 6(1), the dual regime created by the ePrivacy Directive makes consent the dominant interaction in practice: Article 5(3) ePrivacy requires consent to store or access information on the user's devices (e.g., for the purposes of cookies), and downstream processing under the GDPR frequently relies on the consent thereby obtained. The result, in the digital environment, is hundreds of consent requests per month. The cognitive burden of the broader notice-and-choice model has been studied by many scholars, including McDonald and Cranor[14] who estimated that reading every privacy policy a typical internet user encounters would have required approximately 76 working days per year– a figure that, given the proliferation of connected services since 2008, is almost certainly an underestimate today. The predictable result is consent fatigue: users faced with repeated, cognitively demanding choices develop heuristics that bypass deliberation entirely. Empirical studies of cookie consent banners consistently find that most users click “accept all” within seconds, without reading the presented options or understanding what they are agreeing to. Acquisti, Brandimarte and Loewenstein[15] document that privacy decisions are characterised by uncertainty, context-dependence, and malleability – properties that are precisely the conditions under which decision fatigue is most severe. The regulatory implication is significant: a framework that relies on repeated, individual consent as its primary legitimating mechanism generates outcomes that are, in aggregate, indistinguishable from a regime of no consent at all. The volume of choices does not produce informed decisions; it produces mechanical acceptance. This observation connects directly to the Digital Omnibus proposals on consent management (Articles 88a and 88b), which seek to replace the site-by-site consent model with more centralised mechanisms. However, among the competitive implications of Article 88b’s browser-level consent signals there is the risk that automated consent mechanisms may inadvertently concentrate (even more) market power in the hands of browser vendors and consent management platforms. A second risk is one of effectiveness rather than market structure. A browser-level signal delivers privacy only if recipients are obliged to honour it: the failure of Do Not Track, which has been widely ignored for want of any duty to comply, illustrates how an unenforced signal can generate a false sense of protection which can be more corrosive than no signal at all.[16] The Global Privacy Control succeeds only where, as in California, it carries the force of law.[17] Default-blocking by browsers (already practised by Safari and Firefox) can improve outcomes,

[14] McDonald, A. M., & Cranor, L. F. (2008). The Cost of Reading Privacy Policies. I/S: A Journal of Law and Policy for the Information Society, 4(3), 543–568.
[15] Acquisti et al. (2015).
[16] See W3C, *Tracking Preference Expression (DNT)*, W3C Working Group Note, last published 17 January 2019. Available at https://www.w3.org/TR/tracking-dnt/.
[17] See California Department of Justice, Office of the Attorney General, *Attorney General Bonta Announces Settlement with Sephora as Part of Ongoing Enforcement of California Consumer Privacy Act*, press release, 24 August 2022. Available at https://oag.ca.gov/news/press-releases/attorney-general-bonta-announces-settlement-sephora-part-ongoing-enforcement. Sephora agreed to pay USD 1.2 million for,

but it sharpens rather than resolves the competitive concern: it vests further gatekeeping discretion in browser vendors, which can make privacy and market-concentration objectives pull against one another, a tension we examine in Part II, Section 4.

2.1.2 Behavioural biases

Even in the absence of consent fatigue – that is, even in settings where the volume of decisions is manageable – systematic cognitive biases compromise the quality of privacy choices. These are not idiosyncratic failures of individual judgment; they are structural features of human cognition that a choice-centric regulatory framework must confront. *Present bias and hyperbolic discounting* lead individuals to underweight future privacy costs relative to immediate benefits. The decision to share personal data typically yields an immediate reward (access to a service, a personalised recommendation, a social interaction) while the privacy risks – data breaches, profiling, discrimination – materialise, if at all, in a distant and uncertain future. Acquisti, Brandimarte and Loewenstein[18] identify this temporal asymmetry as a key driver of the gap between stated privacy preferences and revealed behaviour. The *default and status quo bias* compounds this effect: users disproportionately accept pre-selected options regardless of their content. In the privacy context, where defaults are typically set to permit processing, this means that the technical configuration of consent interfaces – not the informed judgment of individuals – determines the effective consent rate[19]. *Framing effects* further distort choices: Adjerid et al.[20] demonstrate that the presentation of identical privacy trade-offs – through changes in wording, ordering, or visual emphasis – significantly alters user behaviour, even when the underlying options and their consequences are held constant. Perhaps most counterintuitively, the *control paradox* documented by Brandimarte, Acquisti and Loewenstein[21] shows that providing users with greater perceived control over their data can increase disclosure rather than reduce it, because the sense of agency generates unwarranted confidence in the safety of sharing. Finally, *social desirability bias* in surveys means that stated privacy preferences consistently overstate the degree of concern users act upon — surveys capture what users believe they should say about privacy, not what they do. The cumulative implication of these findings is that the gap between stated and revealed privacy preferences – the phenomenon traditionally labelled the "privacy paradox" – is not a paradox at all, but the predictable outcome of a regulatory model that presupposes rational, informed decision-making in conditions where neither rationality nor information is reliably available. A regulatory framework that relies primarily on individual consent to legitimise data processing will systematically under-protect users – not because users do not care about privacy, but because the cognitive conditions for meaningful consent are structurally unmet.

2.1.3 The transparency gap

The biases documented above operate in a context of pervasive informational opacity that amplifies their effects. Users lack a reliable mapping between the data they share and the downstream

inter alia, failing to honour opt-out preference signals such as the Global Privacy Control, as required under the California Consumer Privacy Act (Cal. Civ. Code § 1798.135).

[18] Acquisti et al. (2015).

[19] Madrian, B. C., & Shea, D. F. (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. Quarterly Journal of Economics, 116(4), 1149–1187.

[20] Adjerid, I., Acquisti, A., Brandimarte, L., & Loewenstein, G. (2013). Sleights of Privacy: Framing, Disclosures, and the Limits of Transparency. Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS).

[21] Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced Confidences: Privacy and the Control Paradox. Social Psychological and Personality Science, 4(3), 340–347.

consequences of that sharing. This is not a matter of insufficient effort on the part of data subjects; it is a structural feature of data markets in which the value, uses, and risks of personal data are unobservable at the point of collection. Acquisti et al.[22] provide a comprehensive account of how information asymmetries distort privacy valuations. Users cannot observe how data is aggregated across sources, how it is combined with other datasets, how long it is retained, or what inferences are drawn from it. These informational gaps are not incidental inefficiencies that better transparency requirements could remedy; they are inherent to the economics of data processing, where the commercial value of data often depends precisely on uses that the data subject cannot anticipate at the time of collection. The CNIL[23] analysis frames this as a structural market failure: even well-functioning markets require that transacting parties can observe what they are exchanging, and data markets systematically violate this condition.

The emergence of AI systems as major consumers of personal data has deepened this opacity. Once personal data enters an AI training pipeline, it is incorporated into model parameters in ways that make it practically impossible for either the data subject or the controller to trace what specific data contributed to what specific output. The right to explanation, the right to erasure, and the right to rectification – all of which presuppose that the relationship between input data and processing outcomes can be identified – become technically challenging and, in many cases, infeasible. This AI-specific opacity functions as an amplifier: consent fatigue and behavioural biases are most damaging precisely when users operate in an informational vacuum, and the AI training context represents the most extreme version of that vacuum. The regulatory implications of this technical reality are examined in Part II, Sections 4.2 and 4.3.

2.1.4 Demand for personalisation

The analysis so far has focused on the obstacles that prevent users from making informed privacy choices. But there is a complementary dimension that any adequate account of user interests must address: users do not merely tolerate data processing as the cost of accessing services, they actively demand the personalised experiences that data processing enables. The same users who express concern about data collection expect search engines to anticipate their queries, recommendation systems to surface relevant content, navigation services to learn their routes, and AI assistants to adapt to their communication style. This is not hypocrisy; it reflects the fact that personalisation delivers genuine utility that users recognise and value. This creates what might be called a *personalisation tension*: users want services to "know" them without users feeling surveilled. They want relevance without intrusion, adaptation without exposure. The standard regulatory framing – which treats data minimisation as a presumptive good and data processing as a presumptive risk – does not capture this tension, because it assumes that less data processing is always better for users. In practice, the relationship is more complex. In many domains, such as health, education, accessibility and language services, the user's interest in data processing is not merely commercial but substantive: a diagnostic tool trained on richer data produces more reliable assessments; a language model trained on more diverse datasets performs more accurately across languages and cultural contexts. The user's decision to contribute data to these systems is not a sacrifice of privacy in exchange for nothing, but it is a transaction in which the user receives tangible value. This is not to deny that privacy-preserving

[22] Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. Journal of Economic Literature, 54(2), 442–492.
[23] CNIL (2023). The Economic Impact of the GDPR, 5 Years On. Commission Nationale de l'Informatique et des Libertés.

techniques such as federated learning, differential privacy, and synthetic data can decouple model capability from individual exposure; rather, the point is that even these methods presuppose access to rich and representative data, which a minimisation-by-default regime treats as a risk to be curtailed. This observation is analytically crucial for the paper's core hypothesis. In a substantial range of settings, the interests of users and AI developers are not opposed but structurally aligned: users benefit from more capable models, and more capable models require diverse, representative data. The regulatory challenge, then, is not to restrict data processing as such but to create conditions under which data processing serves both the interests of users and the development of AI systems.

2.1.5 AI as a tool for improving data protection

A final dimension of the user's perspective that deserves explicit attention is the potential for AI technologies themselves to improve data protection outcomes. The standard narrative frames AI and data protection as being in tension: AI systems require data, and data protection restricts access to it. But the relationship also runs in the opposite direction. AI-driven tools can automate and improve the very mechanisms through which data protection is implemented, lowering the costs of compliance for firms and improving the quality of protection for users. Several categories of application are already emerging. Automated consent management tools can help users express and enforce their privacy preferences more effectively than the current regime of manual banner interactions — a development that connects directly to the Omnibus proposals on privacy signals (Articles 88a–88b). Privacy-enhancing technologies (PETs), including differential privacy, federated learning, and synthetic data generation, allow data to be used for model training while reducing or eliminating the exposure of individual-level information. These techniques are not theoretical; they are deployed at scale by major technology firms and are increasingly available as standard components of data infrastructure. On the enforcement side, AI tools can assist Data Protection Authorities in monitoring compliance, detecting violations, and prioritising cases — addressing the well-documented capacity constraints that limit the effectiveness of the current supervisory framework.

The relevance of this observation for the paper's argument is twofold. First, it reinforces the conclusion that data protection and AI development are not inherently in conflict: the same technologies that create new privacy challenges also create new tools for addressing them. Second, it suggests that a regulatory framework that enables responsible AI development may, as a side effect, produce better privacy outcomes than one that restricts it – a consideration that the current GDPR framework does not systematically incorporate.

2.2 Firms' perspective

The preceding subsection documented the obstacles that prevent users from making well-informed privacy decisions, while also identifying settings in which user interests and data processing are structurally aligned. This subsection turns to the firms that develop and deploy AI systems and examines how the current regulatory framework affects their capacity to innovate, compete, and produce the AI-driven services that users demand. The evidence reviewed here draws primarily on the empirical economics literature studying the effects of the GDPR since its entry into force in 2018.

2.2.1 Disproportionate SME compliance burden

The most extensively documented effect of the GDPR on firms is the disproportionate compliance burden it imposes on small and medium-sized enterprises. Regulatory compliance involves both fixed and variable cost components: legal analysis of applicable obligations, implementation of technical safeguards, appointment of data protection officers, development of privacy notices, conduct of data protection impact assessments, and management of data subject access requests. The fixed component of these costs — which must be incurred regardless of the volume of data processed or the revenues of the firm — falls proportionately more heavily on smaller enterprises.

The empirical evidence confirms this prediction. Jia, Jin and Wagman[24] document an approximately 26% reduction in the number of monthly EU technology investment deals following the GDPR's implementation, with the decline concentrated among newer ventures and data-intensive business models — precisely the categories of firm for which fixed compliance costs represent the largest share of operating expenses. Frey and Presidente[25] find an average profit decline of approximately 2.1% for European technology firms attributable to GDPR compliance, with smaller firms experiencing steeper declines. Demirer et al.[26] estimate that EU firms store approximately 26% less data than comparable non-EU firms, with compliance costs ranging from $1.7 million to $70 million depending on firm size and data intensity. Yun[27] provides a comprehensive review of 31 empirical studies and concludes that the pattern is consistent: the GDPR has reduced start-up activity, dampened venture investment in data-intensive sectors, and increased market concentration. These findings are consistent with the broader assessment in Decarolis and Firullo[28], which documents how operational compliance costs dominate technical implementation costs and fall disproportionately on smaller firms. The irony is difficult to miss. A regulatory framework designed, in significant part, to constrain the market power of large technology platforms and other large incumbents has, through the mechanism of disproportionate fixed compliance costs, consolidated their competitive position. Large incumbents can absorb compliance overhead as a routine cost of operations; smaller entrants cannot. The result is a regulatory barrier to entry that advantages incumbents with scale to absorb fixed compliance costs—including large technology platforms, but also other large corporates—not only the firms whose data practices originally motivated privacy regulation.

2.2.2 Reduced innovation incentives

The compliance burden documented above has downstream effects on innovation. Higher regulatory costs reduce the expected returns to data-intensive innovation, discouraging European investment in precisely the sectors – AI development, personalised services, data analytics – where Europe's

[24] Jia, J., Jin, G. Z., & Wagman, L. (2021). The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment. Marketing Science, 40(4), 661–684.
[25] Frey, C. B., & Presidente, G. (2024). Privacy Regulation and Firm Performance: Estimating the GDPR Effect Globally. Economic Inquiry, 62(3), 1074–1089.
[26] Demirer, M., Jiménez Hernández, D. J., Li, D., & Peng, S. (2024). Data, Privacy Laws and Firm Production: Evidence from the GDPR. NBER Working Paper No. 32146.
[27] Yun, J. M. (2024). A Report Card on the Impact of Europe's Privacy Regulation (GDPR) on Digital Markets. George Mason Law Review Forum, 31, 104–124.
[28] Decarolis, F., & Firullo, C. (2025). Unlocking Growth: Exploring the Economic Impact of GDPR for Tomorrow's Europe. Report prepared for Amazon.

competitive position is most fragile. Jia et al.[29] and the updated analysis in Jia et al.[30] document a significant pullback in venture capital investment in EU data-intensive start-ups following the GDPR's entry into force, with the effect particularly pronounced for foreign investors. The evidence is consistent with the view that the GDPR operates not only as a direct compliance cost, but also—alongside other structural features of the EU economic environment—as a factor that can shape investors' perception of regulatory risk. Survey and market evidence regularly cite, in addition to data-protection rules, the fragmentation of capital and financial markets, broader regulatory complexity, and scale disadvantages for new entrants as constraints on investment in EU-based, data-intensive ventures. In that sense, attributing investment shortfalls to the GDPR alone would overstate its explanatory power; it is more accurately described as one element in a wider competitiveness and risk narrative[31]. Johnson[32] provides a comprehensive survey of the economic research on privacy regulation and innovation, concluding that the effects are not uniformly negative but are context-dependent: privacy regulation can redirect innovation toward compliance-oriented products (privacy-by-design tools, consent management platforms) while reducing investment in data-intensive applications. The CNIL[33] offers a more nuanced assessment, suggesting that compliance costs should be understood as an "investment in regulatory quality" that can yield long-term benefits in consumer trust and market stability. Both perspectives are relevant: the question is not whether the GDPR has any innovation-dampening effects, but whether specific provisions – and the Omnibus amendments to those provisions – achieve a proportionate balance between regulatory protection and innovation incentives.

2.2.3 Value of diverse datasets

A distinct but related concern is the role of data diversity in the development of AI systems that perform well, serve diverse populations equitably, and adequately represent the EU's cultural and linguistic heterogeneity. The performance of AI models is a direct function of the breadth, depth, and representativeness of their training data. The risks of training on narrow or biased datasets have been documented across multiple AI modalities. Buolamwini and Gebru[34], in their influential "Gender Shades" study, demonstrated that commercial facial recognition systems exhibited dramatically higher error rates for darker-skinned female subjects – a failure directly attributable to underrepresentation in training data. Bender et al.[35] generalised this concern to large language models, showing that the predominantly anglophone and Western character of the corpora on which LLMs are trained reproduces and amplifies the linguistic, cultural, and ideological biases embedded in those corpora – with consequences that fall disproportionately on populations whose languages and contexts are under-represented in available digital data. The lesson generalises across domains: AI systems trained on narrow or biased datasets will produce outputs that are narrow or biased. The EU faces a particularly acute version of this challenge. Joshi et al.[36] document that of the world's roughly

[29] Jia, J., Jin, G. Z., & Wagman, L. (2020). The Short-Run Effects of GDPR on Technology Venture Investment. SSRN Working Paper.
[30] Jia, J., Jin, G. Z., Leccese, M., & Wagman, L. (2025). How Does Privacy Regulation Affect Transatlantic Venture Investment? Evidence from GDPR. NBER Working Paper No. 33909.
[31] See also Decarolis and Firullo (2025); Draghi (2024).
[32] Johnson, G. (2022). Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond. NBER Working Paper No. 30705.
[33] CNIL (2023).
[34] Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the Conference on Fairness, Accountability and Transparency (FAT*), 77–91.
[35] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 610–623.
[36] Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), 6282–6293.

7,000 languages, only a small minority are well-served by current NLP infrastructure, with European languages other than English systematically under-resourced relative to anglophone benchmarks. With 24 official languages, significant cultural variation across Member States, and relatively smaller volumes of digitised data compared to anglophone markets, the EU's AI ecosystem depends on access to diverse datasets to produce systems that serve its population adequately. The AI Act itself, in Article 10 of Regulation (EU) 2024/1689, mandates that training, validation and testing data sets for high-risk AI systems must meet quality criteria including representativeness – but achieving representativeness requires data access, and data access is precisely what privacy regulation constrains. Overly restrictive rules on data processing may paradoxically harm the populations they are designed to protect, by ensuring that AI systems are less accurate, less fair, and less adapted to the diversity of their intended users. This regulatory tension is one that the Digital Omnibus attempts, in part, to address through the new legitimate-interest basis for AI training (Article 88c GDPR) and the special-category derogation for bias detection and mitigation (Article 9 amendments and Article 4a Omnibus on AI; see Section 4.2.3).

2.2.4 AI as compliance tool

Just as AI technologies can improve privacy outcomes for users (Section 2.1.5), they can also lower the costs of regulatory compliance for firms. Automated data mapping tools can identify and classify personal data across an organisation's systems, reducing the manual effort required for data protection impact assessments. AI-driven compliance monitoring can detect potential violations in real time, enabling proactive remediation rather than reactive response to supervisory action. Automated consent management, intelligent data classification, and anomaly detection in data processing activities all represent applications where AI reduces the cost of doing what the GDPR already requires. This creates the potential for a virtuous cycle: a regulatory environment that enables responsible AI development also produces the tools that lower the cost of regulatory compliance, which in turn frees resources for productive innovation. Conversely, a regulatory environment that restricts AI development may deny firms access to the very tools that would help them comply more efficiently and more effectively. The Omnibus provisions examined in Part II — particularly the legitimate interest basis for AI training and the streamlined DPIA framework — should be evaluated in part on whether they enable or obstruct this dynamic.

2.3. The dual framing as foundation for Part II

The evidence reviewed in this section points to a conclusion that departs from the standard framing of the policy debate. Data protection and AI-driven innovation are not, in general, opposed interests requiring a zero-sum trade-off. In a substantial range of settings – personalisation, bias mitigation, linguistic diversity, automated compliance, privacy-enhancing technologies – the interests of users and developers are structurally aligned. Users benefit from more capable AI systems, and developers need diverse, high-quality data to build them. Where interests genuinely diverge – as in the use of personal data for purposes that users neither anticipate nor benefit from – the case for regulatory intervention is strong. The challenge for the regulatory framework, then, is not to choose between privacy and innovation but to distinguish between scenarios of alignment and scenarios of divergence. So far, data protection authorities have shown relatively little willingness when enforcing GDPR to strike a balance between privacy and innovation or, where relevant, between privacy and other fundamental rights. Too often, GDPR rules have been implemented with an absolutist focus on the

fundamental right to privacy; hence, where the GDPR provisions conflict with other rules, values or principles, little effort has been made to allow controllers to propose alternative privacy-enhancing solutions that achieve a more proportionate balance. For example, in the past we have seen DPAs promoting an undifferentiated reliance on consent as the primary or preferred legal basis for processing data – albeit recently with some more nuances. Individual consent has been applied to settings where user and developer interests converge (and where consent is both unnecessary and ineffective) and to settings where they diverge (and where consent, even if meaningful, may be insufficient). The result is a framework that is simultaneously over-inclusive, in that it burdens processing activities that serve user interests, and under-protective, in that it relies on a mechanism whose cognitive preconditions are systematically unmet. Each provision of the Digital Omnibus examined in Part II should be evaluated against this dual standard. Does it help where interests converge – by removing unnecessary friction, enabling beneficial data use, and lowering compliance costs without reducing protection? Does it protect where interests diverge – by ensuring that data subjects retain meaningful control over uses they have not authorised and from which they do not benefit? Or does it apply a one-size-fits-all approach that fails on both counts? The substantive analysis that follows is organised around these questions.

3. The background: EU digital regulation at the intersection of data and AI

The right to privacy and the right to data protection, which respectively concerns the intimacy of the individual's private sphere and the respect of safeguards governing the processing of personal data, enjoy a higher level of protection in the European Union compared to other jurisdictions. Scholars[37] have argued that privacy is "Europe's First Amendment" to emphasise the comparatively higher consideration these fundamental rights have received in Europe than in other legal systems. It does not come as a surprise that the Treaty on the Functioning of the European Union ("TFEU") mentions the right to data protection in its Article 16, which reflects the key value attached to personal data from both a market and a human rights perspective. Indeed, the right to data protection has served a pivotal role in the harmonisation of the internal market. At the same time, its protection in the EU legal system has been driven by the European integration process. It is not coincidental that the implementation of Directive 95/46/EC became closely linked to participation in the Schengen acquis. Since Schengen cooperation depends on extensive cross-border data sharing, compliance with EU data protection standards was regarded as a key prerequisite. As a matter of fact, various Member States still lacked comprehensive personal data protection legislation at the time Directive 95/46/CE was adopted. Equally remarkable is that Directive 95/46/EC reflected a twofold logic: while seeking to protect individuals' fundamental rights, it was also designed to facilitate the free movement of personal data within the internal market. The recognition of data protection as a distinct autonomous fundamental right would emerge more clearly only with the EU Charter.

Directive 95/46/EC was a technology-neutral piece of legislation, which could not consider the magnitude of the digital transition that emerged a few years later, shaping data collection and processing practices in both the public and private sectors. This is perhaps the reason why Directive 95/46/EC, as well as the relevant implementing laws adopted by Member States, soon became ill-suited to the reality of a digital economy. Building upon the Directive's foundations, the GDPR introduced a risk-based and accountability-oriented approach, providing more adequate responses to the challenges posed by the digital transition. While remaining technology-neutral, the GDPR reflects a more sophisticated understanding of processing activities, particularly in light of their predominantly automated nature and the impact of digital technologies.

In the meantime, the right to data protection formally gained a specific and autonomous recognition in the Charter of the Fundamental Rights of the European Union under Article 8, while Article 7 protects the right to respect for private and family life and largely mirrors Article 8 of the European Convention of Human Rights. With the introduction of Article 16 TFEU, the European Union codified a specific competence to legislate on the matter, marking a shift from a predominantly internal market rationale towards the constitutional recognition of data protection as an autonomous fundamental right. Article 16 thus became the legal basis for the adoption of the GDPR in 2016. The entry into force of the new regulation marked a departure from a scenario of legal fragmentation (depending on the existence of 28 different national laws implementing Directive 95/46/EC). However, the making of the GDPR also reflected some key achievements in the case law of the Court of Justice, such as the landmark *Google Spain* judgment concerning the territorial scope of application of EU data protection

[37] Petkova B. (2019). Privacy as Europe's first Amendment. European Law Journal, 25(2), 140-154.

law and the enforcement of the right to erasure[38] and the *Schrems I* judgment[39], which invalidated the European Commission's 2005/520/EC adequacy decision on the transfer of personal data to the United States. The Court of Justice has significantly contributed to interpreting EU data protection law also in the aftermath of the adoption of the GDPR, consistently interpreting and applying the GDPR in light of the technological developments.

Against this background, however, EU data protection law, and most notably the GDPR, is sometimes perceived as imposing constraints on data-driven innovation and data flows while failing to effectively empower individuals. At times, the GDPR is said to fall short in providing legal certainty considering unprecedented technological developments. New challenges emerge because of the evolution of business models premised on data collection and processing practices, such as the legal feasibility of pay-or-consent schemes[40]. Calls for revisiting the GDPR and proposed improvements are increasingly frequent[41]. Some commentators interpret the "reopening" of the GDPR as a threat, others see it as an opportunity[42]. The large-scale implementation of AI systems has further stimulated an assessment of the adequacy of the GDPR categories with respect to various problems, spanning the definition of the legal bases for AI model training, the enforcement of data subject rights, the establishment of proper oversight mechanisms, among others.

The rise of calls to boost European competitiveness seems to have finally revamped the debate on whether the GDPR is still a good fit for AI technologies.

The intervening decade has also altered the global context within which EU data protection law operates. The Union's share of global GDP has declined from approximately 22% in 2015 to roughly 14% in 2024[43], while other major economies – notably China, where institutional-logics scholars have documented the emergence of a tri-polar fragmentation of cross-border data rules[44] and proposed alternative international governance arrangements such as a "World Data Organization" designed to address the perceived institutional gap[45] – are pursuing rule-making as an instrument of geopolitical influence rather than passively importing EU standards[46]. The assumption that the so-called Brussels effect[47] will continue to operate as the de facto global benchmark is no longer self-evident. Against this background, the Digital Omnibus may also be read as a response to concerns regarding the declining economic weight of the Union and the resulting pressure on the continued global influence of its regulatory model – a reading that complements the competitiveness-driven framing developed in the Draghi and Letta reports and now embedded in the package. Not coincidentally, the Digital Omnibus pursues the goal of simplification as a way to reduce the burden for controllers and thus unlock new opportunities in the EU digital market. It does so within the framework drawn by the Draghi report and the Letta report. This is not a neutral factor in the evaluation of the measures under the umbrella of the Digital Omnibus: the proposed amendments to an array of EU law acts, and most

[38] ECJ, Case C-131/12, *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, judgment of 13 May 2014.
[39] ECJ, Case C-362/14, *Maximillian Schrems v Data Protection Commissioner*, judgment of 6 October 2015.
[40] ECJ, Case C-252/21, *Meta Platforms Inc and Others v Bundeskartellamt,* judgment of 4 July 2023.
[41] Kloza, D., Dreschsler L., & Fernandes, E. (2026). If it ain't broke, don't fix it? Ten improvements for the upcoming tenth anniversary of the General Data Protection Regulation. Computer Law & Security Review, 60, 106251, 1-31.
[42] Voss A. (2025). We should revise the GDPR to unlock Europe's digital future. CEPS.
[43] Eurostat (2025). EU economy in the world. Available at https://ec.europa.eu/eurostat/web/products-eurostat-news/w/wdn-20250326-1.
[44] Rong, K. et al. (2025). Cross-border data transfer: patterns and discrepancies. Journal of International Business Policy, 8, 10-32.
[45] Rong, K. et al. (2026). World Data Organization: Filling the Institutional Gap in Cross-Border Data Governance. Journal of Digital Economy, forthcoming.
[46] Rong et al. (2026); Rong et al. (2025).
[47] Bradford (2020).

importantly to the GDPR, are driven by the objective of promoting innovation by reducing red tape for business operators. This represents a notable normative shift, as the GDPR does not articulate innovation among its explicit regulatory objectives. Rather, its two operative pillars, as reflected in Article 1(2)–(3), are the protection of natural persons and the free movement of personal data. The term "innovation" does not appear in the operative text, although GDPR was not agnostic to innovation and was supposed to foster it. By contrast, the Digital Omnibus explicitly mentions "innovation" thirty-three times across its operative text and recitals, framing it explicitly as a regulatory aim. The shift is normative rather than interpretive, and the present analysis treats the Digital Omnibus as introducing a new objective rather than as merely operationalising one already embedded in the GDPR. Concerns related to a possible weakening of fundamental rights protection are understandable, but the recent stances of the EU institutions seem consistent with the political goal of boosting European competitiveness. While the EU made data protection a flagship of European constitutionalism over the past two decades, recent initiatives suggest that innovation, competitiveness and technological sovereignty increasingly occupy a more prominent place on the Union's agenda. These moves are consistent with the view that economic integration is a foundational factor for the EU and one of the key justifications of its existence.

The European Union's change of posture is also visible in the shift from the European Strategy for Data, which led to the adoption of the Data Act and the Data Governance Act, to a European Data Union Strategy, which was announced along with the Digital Omnibus in November 2025. The goal of the Data Union Strategy is to increase the availability of data for AI development and to simplify EU data rules, also with a view to strengthening the EU's global position on international data flows. It reflects the awareness of the opportunities unlocked by AI technologies in both the public and private sectors and thus sets as a priority the need for high-quality data to train AI models in Europe. Developing better AI solutions is supposed to help especially SMEs to compete in a global economy that the Strategy defines as increasingly shaped by AI. If the EU wishes to deliver on these promises, however, some inherent tensions between the existing data protection framework and the peculiarities of AI must be acknowledged and addressed. Whether these tensions justify a recalibration of existing safeguards remains a contested question.

This is what the Digital Omnibus aims to do, at least partially. A first instance of conflict is between the principle of data minimisation enshrined in the GDPR and the data-intensive nature of AI technologies, which would otherwise follow a "data maximisation" rationale[48]. Transparency also requires technical conditions that may not be easy to meet in the specific context of AI, given the impact of the "black box" problem and the inherent opacity of these systems. The principles of lawfulness and fairness are in turn difficult to interpret in light of the characteristics of AI, particularly when it comes to determining the legal basis for processing personal data for AI model training. Finally, the exercise of data subject rights can prove difficult to reconcile with the technical features of AI. Whereas these principles directly capture some tensions between EU data protection law and AI systems, there are other aspects that indirectly deserve attention given their ability to impact data collection practices. In this respect, the Digital Omnibus considers the problem of consent fatigue and the reasons why individuals face increasing challenges in retaining control over their personal information and how automation can help safeguard data subjects while also creating more legal certainty for controllers.

[48] Zornetta, A., & Cofone, I. (2023). Artificial Intelligence and the Right to Privacy. In Temperman, J., & Quintavalla, A. (eds). Artificial Intelligence and Human Rights. Oxford University Press, 121-135.

Part II

4. The Digital Omnibus: a legal and law-and-economics analysis

In response to emerging calls for a recalibration of the EU digital rulebook to stimulate business dynamism and foster digital competitiveness, the European Commission released a new Digital Package in November 2025[49]. In the European Commission's view, the package "is designed to help EU businesses innovate, scale, and save on administrative costs" and directly responds to the recommendations in the Draghi Report, among others, "to boost productivity through innovation in digital" including by "addressing barriers to regulatory compliance" and "boosting access to high-quality data across Europe". A crucial component of the package is what has become commonly known as the "Digital Omnibus", which amends personal and non-personal data and cybersecurity rules and certain elements of the AI Act.[50] The Digital Omnibus consists of two proposals for regulations: the first (COM/2025/837) amending the GDPR, the Data Act, the Data Governance Act as well as the NIS2 Directive and the ePrivacy Directive, which is commonly referred to simply as the "Digital Omnibus"; the second (COM/2025/836) introducing targeted amendments to the AI Act, which is described as the "Digital Omnibus on AI". Both are currently under negotiation between the European Parliament and the Council under ordinary legislative procedures. Although presented as part of the same package, the two proposals have followed different legislative trajectories. In particular, the proposal amending the AI Act reached a provisional political agreement in trilogue on 7 May 2026[51], which was subsequently endorsed by the Parliament's IMCO and LIBE Committees on 2 June 2026[52]. By contrast, negotiations on the broader Digital Omnibus proposal remain ongoing.

In the European Commission's own framing, the package is designed to reduce unnecessary administrative burdens, lower compliance costs – particularly for SMEs – and create a regulatory environment that fosters innovation while preserving the EU's high standards of fundamental rights protection.

The changes proposed in the two proposals reflect some of the most debated issues at the intersection of data protection and AI regulation, which have already arisen before courts and regulators, and whose relevance extends beyond the specific legislative trajectory of the Omnibus itself. Some of the proposed modifications do not directly concern AI-related developments; however, they are overall likely to make an impact on the circulation of personal data in a way that may enable unlocking new

[49] European Commission (2025). Simpler EU digital rules and new digital wallets to save billions for businesses and boost innovation. Press Release. Available at: https://digital-strategy.ec.europa.eu/en/news/simpler-eu-digital-rules-and-new-digital-wallets-save-billions-businesses-and-boost-innovation.

[50] European Commission, Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2022/868 and (EU) 2023/2854 as regards the simplification of the digital legislative framework, COM(2025) 837 final; and Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) 2024/1689 as regards the simplification of the implementation of harmonised rules on artificial intelligence, COM(2025) 836 final.

[51] See European Commission (2026). EU agrees to simplify AI rules to boost innovation and ban 'nudification' apps to protect citizens, 7 May; Council of the European Union (2026). Proposal for a regulation of the European Parliament and of the Council as regards the simplification of the implementation of harmonised rules on artificial intelligence, Letter sent to the European Parliament, 13 May.

[52] European Parliament - Committee on the Internal Market and Consumer Protection Committee on Civil Liberties, Justice and Home Affairs (2026). Amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), 2 June.

opportunities in the AI industry. For these reasons, regardless of the outcome of the ongoing negotiations between the European Parliament and the Council, the proposed amendments provide a unique opportunity to reflect upon the compatibility of the existing data protection rules with the goal of promoting innovation in the field of AI, while safeguarding fundamental rights. This exercise is particularly complex. While the right to data protection has gained considerable traction in EU law, also by virtue of the judicial activism of the Court of Justice[53], the Digital Omnibus has sparked a quite significant criticism among scholars for allegedly weakening this right without delivering a proportionate gain in regulatory dynamism for its intended beneficiaries – particularly, SMEs[54]. Also, the very same nature of the Omnibus legislation marks an important limitation for the ambitions of this reform. On the one hand, the European Commission stated that the proposed amendments are "targeted and technical in their nature", "designed to ensure a more efficient implementation of rules" and "not prone to multiple policy options that could meaningfully be tested and compared"; accordingly, it pointed out that they did not need to undergo a full impact assessment. On the other hand, however, the Digital Omnibus seems in its content to have a broad and quite substantial impact on some foundational norms, most notably in data protection law, bearing significant effects for the right to privacy and the right to data protection. Therefore, in such circumstances, carrying out a proper impact assessment would have paved the way for a better understanding of the implications of the proposal, also in terms of cost-benefits analysis[55]. It is therefore inherently difficult to appraise the expected consequences of the Omnibus provisions without access to the underlying evidential basis.

The following subsections provide a substantive analysis of a selection of provisions of the Digital Omnibus and the Digital Omnibus on AI, most of which are currently under negotiation, examining the rationale behind each and assessing whether they deliver the flexibility and legal certainty that a competitive AI market requires.

4.1. Personal data in the age of AI: a contested definition?

Among the most structurally significant proposals contained in the Digital Omnibus is the attempted revision of the definition of personal data under Article 4(1) GDPR. The European Commission proposed adding a new paragraph to that provision, introducing a relative and entity-specific approach to identifiability. The proposed text would have read: "Information relating to a natural person is not necessarily personal data for every other person or entity, merely because another entity can identify that natural person. Information shall not be personal for a given entity where that entity cannot identify the natural person to whom the information relates, considering the means reasonably likely to be used by that entity. Such information does not become personal for that entity merely because a potential subsequent recipient has means reasonably likely to be used to identify the natural person to whom the information relates". A companion provision – the proposed Article

[53] See Brkan, M., & Psychogiopoulou, E. (eds) (2017). Courts, Privacy and Data Protection in the Digital Environment. Edward Elgar; see also Pollicino, O. (2021). Judicial Protection of Fundamental Rights on the Internet. A Road Towards Digital Constitutionalism?. Hart Publishing.
[54] Hofmann, H.C.H. (2026). This is Not Simplification. Verfassungsblog, 3 January. See also Bieker, F. (2026a). European AI FOMO. Verfassungsblog, 29 January.
[55] See also Bassini, M., Maggiolino, M., & de Streel, A. (2025). Better Law-Making and Evaluation for the EU Digital Rulebook. CERRE Report. January.

41a – was meant to empower the European Commission to adopt implementing acts specifying the conditions under which pseudonymised data ceases to constitute personal data for certain entities.

The rationale behind this proposal was twofold. First, the European Commission presented the amendment as a codification of the ECJ's judgment in *EDPS v SRB*[56], in which the Court confirmed that the application of the GDPR depends on the actual ability of the relevant controller to identify the data subject, to be assessed on the basis of the means reasonably likely to be used by that specific controller [57]. In the Court's reasoning[58], the same dataset may therefore constitute personal data for the disclosing controller while falling outside the scope of the GDPR for a recipient that lacks the means to re-identify the individuals concerned. This issue is particularly salient in AI development, where large-scale training datasets are frequently pseudonymised and processed by multiple actors across the value chain. Second, in the European Commission's view, codifying this principle would bring legal certainty to controllers – including AI developers – operating on pseudonymised datasets in complex multi-party processing schemes. For the AI industry, particularly in sectors such as health, financial services, analytics, the proposal held the promise of removing a significant source of uncertainty in model development: whether pseudonymised training data remains within the scope of the GDPR from the perspective of every actor in the chain[59].

The proposal attracted strong opposition[60]. In their Joint Opinion 2/2026[61], adopted on 10 February 2026, the EDPB and EDPS urged co-legislators to reject the amendment in its entirety. First, in the EDPB and EDPS's view, the revision went far beyond what the European Commission had described as a "targeted" modification: as the definition of personal data determines when the fundamental right to data protection applies, any alteration of that concept carries structural consequences for the entire GDPR architecture and, by extension, for the protection afforded to data subjects under Article 8 of the Charter. As noted, the definition of personal data has acquired a quasi-constitutional status in EU law. Modifying it therefore amounts to a substantive constitutional intervention rather than an ordinary legislative refinement[62]. Second, the supervisory authorities contended that the third sentence of the proposed addition — providing that information *does not become* personal data merely because a subsequent recipient has means reasonably likely to be used for identification — was not, in fact, supported by the *SRB* judgment[63]. On the contrary, the EDPB and EDPS argued that the *SRB* judgment reaffirmed the principle that data *may become* personal in nature when made available to any recipient with the technical capacity to identify the data subject, meaning that such data constitute personal data both for the recipient and, indirectly, for the entity making the data

[56] ECJ, Case C-413/23 P, *European Data Protection Supervisor v Single Resolution Board,* judgment of 4 September 2025.
[57] Duarte, T. (2025). Identifiability on Trial: Insights from SRB v EDPS. CiTiP Blog, KU Leuven; Spajić, D. (2023). Anonymous vs. pseudonymous data: the CJEU reaffirms the relative approach to the concept of personal data. CiTiP Blog, KU Leuven.
[58] Chiara, P.G. (2025). The Court of Justice Upholds a Relative Approach to EU Data Protection Law. European Data Protection Law Review, (11)4, 566-572.
[59] Wendehorst, C. (2025). Who is afraid of the Digital Omnibus?. EUCO Debrief IV/2025. Trans European Policy Studies Association. 31-32.
[60] As noted by González Fuster, G. (2026). Caught Between AI and the AI Hype: How the Right to Personal Data Protection was Ambushed. Rivista di Diritti Comparati, VIII, 107-123, at 119, the proposed amendments seem to conflict with the current wording of Recital 26 GDPR, which requires taking into account all the means reasonably likely to be used to identify the natural person (directly or indirectly) either by the controller or *by another person*. See also EDPB (2025). Guidelines 01/2025 on Pseudonymisation, 16 January. See also Ruschemeier, H. (2025). The Omnibus Package of the EU Commission. Verfassungsblog, 17 November.
[61] EDPB-EDPS (2026b). Joint Opinion 2/26 on the Proposal for a Regulation as regards the simplification of the digital legislative framework (Digital Omnibus).
[62] Kramcsák, P.T., & Lazarotto, B. (2026). "Personal Data": More Than a Definition, a Quasi-Constitutional Stake in EU Law in the Era of the Digital Omnibus. European Law Blog.
[63] Bieker, F. (2026b). Schroedinger's data: SRB and the Digital Omnibus. European Law Blog; Mannekens, J. (2026). Personal data in the Digital Omnibus: where are we going?. CiTiP Blog, KU Leuven.

available[64]. A codification selectively extracting only certain elements of the Court's reasoning – without incorporating its contextual and relational safeguards – risked causing internal incoherence within the ECJ's line of identifiability jurisprudence[65]. The EDPB and EDPS further objected to the delegation of authority to the European Commission under the proposed Article 41a, noting that determining what ceases to be personal data following pseudonymisation is a matter of constitutional significance that should not be entrusted to implementing acts.

Opposition to amending the personal data definition has also manifested among co-legislators. While negotiations remain ongoing, maintaining the European Commission's original approach appears extremely unlikely. Council's positions have instead tended to favour guidance — developed by the EDPB rather than a legislative redefinition — on pseudonymisation and identifiability.

The Council's view on this matter seems plausible, as a degree of simplification in this respect may be achieved without revisiting foundational concepts in data protection law through proper guidance from the competent authorities. The shift from the European Commission to the EDPB in this respect reflects a reasonable choice, which would avoid turning this clarification into a de facto exercise of law-making power that could result in unintended consequences.

The Council's preference for guidance rather than legislative redefinition may also be justified on economic grounds. From an economic standpoint, the proposed amendment diverges from the simplification objective the Digital Omnibus articulates. A relative, entity-specific definition of personal data shifts the cost of identifiability assessment from the rule to each controller-recipient interaction, multiplying compliance costs along multi-party processing chains typical of AI development. The distributional incidence is asymmetric: larger controllers absorb iterative assessments as a marginal cost, while SMEs, whose fixed compliance costs already weigh disproportionately under the GDPR (Section 2.2.1), face a new fixed-cost barrier in precisely the markets the Digital Omnibus seeks to support. Net effect on legal certainty would likely be negative: substituting a centralised judicial doctrine (SRB) with an entity-by-entity test converts a question of law into a question of fact, generating evidentiary disputes in every enforcement action. The Council's deletion is therefore consistent not only with the EDPB-EDPS legal critique but also with the economics of compliance: a single, generally applicable definition lowers the fixed costs of legal interpretation by avoiding the duplicated, case-specific effort that a relative standard imposes. These savings accrue disproportionately to SMEs as these costs are largely fixed[66].

This aspect of the Digital Omnibus is instructive for the purposes of this research in two respects. First, it illustrates the limits of simplification strategies that operate through changes to foundational definitional concepts[67]: far from delivering legal certainty, the amendment would have likely increased the interpretive instability that the Digital Omnibus was designed to avoid. Second, it confirms that the tensions between data protection and emerging technologies (such as AI) are not productively resolved by narrowing the material scope of the GDPR but require a more refined recalibration. It is

[64] See EDPB-EDPS (2026b), para. 16.
[65] Bieker (2026b).
[66] See Decarolis & Firullo (2025) on the disproportionate impact of regulatory fixed costs on SMEs.
[67] This is one of the areas in which, as noted by the Meijers Committee in its Comment on the Digital Omnibus Proposal (CM2604, February 2026), an impact assessment – particularly with regard to fundamental rights – would have been necessary as a key instrument of better regulation.

to those changes — and to the specific Omnibus provisions that address them — that the following sections turn.

4.2 AI model training

AI systems need more and higher quality data to perform better: the more the training of the underlying models is structured, the more patterns and correlations will be easy to detect and learn. Thus, data play a crucial role for the development of AI technologies, and this conclusion already signals an inherent tension between the reliance of AI on personal data and the principle of data minimisation behind the GDPR. The rise of a European market for AI has therefore to deal with the existing strictures and constraints posed by EU data protection law, and a key problem at the intersection of the two lies in the definition of the suitable legal bases for the processing of personal data for AI model training[68].

Under Article 6 GDPR, the lawfulness of processing of personal data (in accordance with the relevant principle enshrined in its Article 5) is conditional on a set of legal grounds, which include, among others, the data subject consent, the legitimate interest of the controller or the performance of a contractual agreement. None of these conditions are per se incompatible with the processing of personal data for training purposes. However, the lawfulness of the use of these legal grounds depends on the specific circumstances in which the processing takes place. The processing of personal data for training purposes can occur in both the system development phase (the design, training and validation of an AI models) and the system deployment phase (the use of an AI system in operational settings under the authority of a deployer) in the sense established by the AI Act.[69] In the development phase, processing activities may occur in the collection of personal data (for example, via web scraping), in their pre-processing and in the actual training operations. In the deployment phase processing activities may occur in the input or output but also in the training of AI models based on the prompts input by users.

Acknowledging the inherent complexity of the processing activities taking place in the AI lifecycle is therefore a key step to better determine the relevance of AI model training and the problems it poses from a data protection standpoint. Such a sophisticated framework also signals that processing activities may fall under the responsibility of different actors, including the providers that develop AI models, those who may develop models on their behalf and even the deployers, namely those subjects using AI systems under their authority in a professional context, as defined in the AI Act.[70] Before turning to the specific Digital Omnibus proposals, it is useful to examine how supervisory authorities at national and European level have already addressed the question of legal bases for AI model training, in the absence of dedicated legislative provisions.

The CNIL guidelines for the legal bases for AI training. The French *Commission Nationale de l'Informatique et des Libertés* (CNIL) has been among the most active national regulators in clarifying

[68] For a comparative overview, Li, W., Zhang, Y., Zheng, Q., & Li, A. (2026). How the Legal Basis for AI Training is Framed in Data Protection Guidelines and Interventions: Comparative Perspectives and the Prospect of Global Convergence. International Data Privacy Law, 1-22.
[69] The terms "development" and "deployment" are used in the sense established by Article 3 of Regulation (EU) 2024/1689 (AI Act): development covers the design, training, and validation of an AI system prior to placing on the market; deployment covers the use of an AI system in operational settings under the authority of a deployer.
[70] Article 3 of the AI Act distinguishes "providers" (entities that develop or commission the development of an AI system and place it on the market — e.g., OpenAI for GPT-4) from "deployers" (entities that use the AI system under their own authority in a professional capacity — e.g., a hospital using an AI diagnostic tool, or a bank integrating a model such as GPT-4 into its customer service).

how the GDPR applies to the development of AI systems. Following successive public consultations, the CNIL consolidated its position in a set of practical "how-to sheets"[71]. On the question of legal basis, the CNIL confirms that, where personal data are collected without direct contact with data subjects — as is typically the case when developers reuse open datasets or resort to web scraping — the legitimate interest under Article 6(1)(f) GDPR will generally be the most appropriate basis, which is also one of the most commonly relied-upon legal bases for AI development by private entities.[72] Importantly, reliance on that basis remains conditioned on the three-part test as interpreted by the Court of Justice from the *Rigas* case onwards.[73]

- the interest pursued must be lawful and precisely defined;
- the processing must be genuinely necessary and consistent with data minimisation; and
- a balancing exercise must establish that the controller's interest is not overridden by the rights and reasonable expectations of data subjects.

The CNIL supplements this framework with a dedicated focus sheet on web scraping, which is not prohibited as such but must be accompanied by specific safeguards, including the prior definition of collection criteria, the exclusion of particularly intrusive sources, and measures limiting the retention of irrelevant data.[74]

The Italian DPA's enforcement trajectory. The lawfulness of processing activities for AI model training already fell under the scrutiny of supervisory authorities at national and European level. The Italian Data Protection Authority (or "*Garante*") broke the ice in its "infamous" temporary ban on OpenAI[75], which was eventually lifted[76]. The *Garante* subsequently imposed a 15 million euro fine on OpenAI[77], then annulled by the Court of Rome on 18 March 2026[78]. In its original decision, the *Garante* found that the processing of personal data by OpenAI for the training of ChatGPT's underlying large language models occurring in the operation of the service and concerning the data input by users in their prompts and conversations could not be justified on the basis of the performance of a contract. On the contrary, such processing activities could rely on either the legitimate interest of the controller or, implicitly, the consent of the data subjects – which is clearly impracticable by default from a business perspective. The *Garante* reaffirmed this orientation in its decision on the Replika chatbot[79], where it did not dispute the lawfulness of the legal ground in question per se but found that the controller had failed to properly substantiate the legitimate interest assessment. On 30 January 2025, the *Garante* imposed a definitive limitation on the processing of Italian users' personal data by DeepSeek[80], the Chinese AI provider, citing concerns about transparency, the absence of an adequate legal basis under the GDPR, and the storage of personal data in China without the safeguards required under Chapter

[71] CNIL, *AI system development: the CNIL's recommendations to comply with the GDPR*, 5 January 2026. Available at https://www.cnil.fr/en/ai-system-development-cnils-recommendations-to-comply-gdpr.
[72] CNIL, AI how-to sheets, Step 3 ("Define the legal basis") and focus sheet 8 bis, Relying on the legal basis of legitimate interests to develop an AI system. Available at https://www.cnil.fr/en/relying-legal-basis-legitimate-interests-develop-ai-system.
[73] ECJ, Case C-13/16, *Rigas,* judgment of 4 May 2017.
[74] CNIL, focus sheet on the safeguards applicable to data collection by web scraping (legal basis of legitimate interest).
[75] Italian DPA, decision of 30 March 2023, doc. no. 9870832.
[76] Italian DPA, decision of 11 April 2023, doc. no. 9874702.
[77] Italian DPA, decision No. 755 of 2 November 2024, doc. no. 10085455.
[78] Court of Rome, 18 March 2026, no. 4785. See also Reuters (2026). Italian court scraps 15-million-euro privacy watchdog fine on ChatGPT-maker OpenAI. Available at https://www.reuters.com/technology/italian-court-scraps-15-million-euro-privacy-watchdog-fine-chatgpt-maker-openai-2026-03-19/.
[79] Italian DPA, decision of 10 April 2025, doc. no. 10130115.
[80] Italian DPA, decision of 30 January 2025, doc no. 10098477.

V GDPR. The DeepSeek case extends the regulatory scrutiny initially developed in the OpenAI proceedings to a non-EU controller, signalling the willingness of supervisory authorities to apply data protection requirements to AI training practices regardless of the controller's geographic location[81]. Similarly to the Italian *Garante* and the French CNIL, other data protection authorities, such as the Belgian one[82], have taken action to provide guidance to business operators regarding decisions concerning the development and deployment of AI models and systems. However, enforcement initiatives are still relatively limited at present.

The EDPB position. At the European level, the European Data Protection Board (EDPB) confirmed the validity of the recourse to legitimate interest as a legal basis in both its Opinion 28/2024 and the Report of the work undertaken by the ChatGPT Taskforce of 23 May 2024[83]. The convergence of national and European supervisory authorities on legitimate interest as the appropriate legal basis is therefore an established feature of the regulatory landscape that the Digital Omnibus enters.

The legitimate-interest framework. The convergence on legitimate interest is, however, conditional on its proper application. The key point of this discussion is not the use of legitimate interest as such, but the conditions that render its use lawful in the development and operation of AI systems and models. Consistent with the risk-based approach behind the GDPR and the accountability principle that informs EU data protection law, legitimate interest should not be interpreted as a trump card but as a legal basis that strongly relies on — and thereby incentivises — controller accountability[84]. As established by the Court of Justice in the *Rigas* case, a three-step test must be carried out: (i) the existence of a legitimate interest; (ii) the necessity of the processing activity; and (iii) its proportionality, taking into account the rights and freedoms of data subjects, which cannot be overridden. The Court has further elaborated upon these conditions and their significance within the three-step test, providing clarification that defines the actual scope of permissible use of the legitimate interest[85].

Against this background, the Digital Omnibus proposal aims to introduce two provisions to regulate the processing of personal data (including special categories) in the context of the development and operation of AI. Furthermore, the Digital Omnibus on AI also addresses the processing of special categories of personal data for the purposes of bias detection and mitigation, in the attempt to rephrase and revisit the content of the debiasing exception currently enshrined in Article 10(5) AI Act. All of them are interesting for a plurality of reasons and merit careful and specific consideration.

[81] Packin, N.G. (2025). A Deep-See On DeepSeek: How Italy's Ban Might Shape AI Oversight. Forbes. Available at https://www.forbes.com/sites/nizangpackin/2025/01/31/a-deep-see-on-deepseek-how-italys-ban-might-shape-ai-oversight.
[82] Gegevensbeschermingsautoriteit (2026). L'impact de l'intelligence artificielle sur la vie privée. Série "IA & Protection des données".
[83] EDPB (2024b). Report of the work undertaken by the ChatGPT Taskforce.
[84] EDPB (2024c). Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR, 8 October 2024. See also Article 29 Data Protection Working Party (2014). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9 April. 844/14/EN WP 217. For an in-depth analysis on the role of legitimate interest and the balancing of its use with rights and freedoms of data subjects see Kamara, I., & de Hert, P. (2018). Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach. In: Selinger, E., Polonetsky, J., & Tene, T. (eds). Cambridge Handbook of Consumer Privacy, Cambridge University Press, 321-352.
[85] More recently, for example, the Court of Justice has found that even a purely commercial interest can amount to a legitimate interest pursuant to Article 6(1)(f) GDPR: see ECJ, Case C-621/22, *Koninklijke Nederlandse Lawn Tennisbond v Autoriteit Persoonsgegevens*, judgment of 4 October 2024 and, for a comment, Bassini, M. (2025). Commercial, but legitimate interest. European Data Protection Law Review, 11(1), 116-122. For an overview of the main cases regarding Article 6(1)(f) see Kotschy, W. (2020). Article 6 Lawfulness of processing. In: Kuner, C., Bygrave, L., & Docksey, C. (eds). The EU General Data Protection Regulation (GDPR): A Commentary. Oxford University Press, 321-344.

4.2.1 Codifying legitimate interest as the legal basis for AI training?

With the proposed introduction of Article 88c GDPR, the Digital Omnibus aims to specifically regulate the processing of personal data for the development and deployment of AI, at first shifting away from the technology-neutral approach behind the GDPR and codifying a dedicated legal basis tailored to a specific technology. In practice, however, such legal basis has already been endorsed by the EDPB in its Opinion 28/2024, and thus the impact in terms of moving away from technology-neutrality remains limited.

Specifically, the first paragraph of the provision establishes that where the processing of personal data is necessary for the interests of the controller in developing or operating an AI system (as defined in Article 3, point (1), of the AI Act) or an AI model, such processing may rely on legitimate interest within the meaning of Article 6(1)(f) GDPR, "where appropriate". This legal basis is excluded (i.) where other Union or national laws explicitly require consent or (ii.) where the interests of the controller are overridden by the interests, fundamental rights and freedoms of the data subject, with intensified protection reserved for children.

The same provision offers some operational guidelines, establishing that the processing must be subject to appropriate organisational and technical measures as well as safeguards for the rights and freedoms of data subjects, so as to ensure (i.) compliance with the data minimisation principle during the selection of sources and the training and testing of AI systems or models, (ii.) protection against disclosure of residually retained data in AI systems or models, (iii.) enhanced transparency to data subjects, and (iv.) provide the latter with an unconditional right to object to the processing of their personal data.

While the European Commission has acknowledged the strategic value of personal data for the development of advanced AI technologies, there are some ambiguities in the current text of the provision that should be addressed and resolved to render this amendment an effective contribution to simplification, notwithstanding the EDPB and EDPS's view that the provision is essentially unnecessary[86].

First of all, the wording "where appropriate", currently embodied in Article 88c, is likely to increase rather than reduce legal uncertainty, as the EDPB and EDPS themselves have noted[87]. It is unclear whether the provision would thereby depart from the well-established three-prong legitimate interest test by introducing an additional condition, whose meaning and practical implications remain unspecified. More critically, however, the provision leaves room for EU law or Member States' laws to derogate from the legitimate interest basis and require consent. In particular, allowing Member States to reintroduce a consent requirement on the basis of their national law would not only create the potential for circumventing the legitimate interest ground, but also for divergent approaches, ultimately increasing fragmentation and undermining the objective of simplification.

Such ambiguity carries a measurable (economic) cost. Legal uncertainty functions as a regulatory risk premium that firms must price into their compliance decisions. For large incumbents with dedicated

[86] EDPB-EDPS (2026b), para. 39.
[87] Ibid, para. 41.

legal teams, this premium is marginal; for SMEs, it may be determinative, tipping the cost-benefit analysis against entering data-intensive markets altogether. As documented in Section 2.2.1, the GDPR's fixed compliance costs already fall disproportionately on smaller firms[88]. A provision designed to simplify the legal framework for AI training that introduces new sources of interpretive discretion risks compounding rather than alleviating this burden. By leaving room for divergent national requirements, the provision risks reintroducing precisely the cross-border fragmentation that the Omnibus seeks to eliminate, compounding the legal-uncertainty premium discussed above.

A second aspect that co-legislators should ideally revisit is the absence of any reference to the legitimate interests of third parties, which renders the legitimate interest in the development and operation of AI systems and models comparatively weaker than the "ordinary" legitimate interest ground under Article 6(1)(f). The proposal offers no apparent justification for excluding, as a matter of principle, the legitimate interests of third parties from the balancing exercise. As noted above with respect to the wording "where appropriate", if adopted in the current form, the provision would risk circumscribing rather than facilitating reliance on the legitimate interest as a legal basis for AI training. In addition to the above, a more general point concerns the balancing test. Weighing the controller's interests against the rights and freedoms of data subjects remains a cornerstone of the legitimate interest ground. As the European Law Institute report on the Digital Omnibus has observed[89], however, the proposal does not fully resolve a practical difficulty that is particularly relevant in the AI context, namely the challenge of conducting a balancing test for potentially vast amounts of training data. How should its conditions be assessed in the context of AI development and deployment? In this respect, the interaction between the GDPR and the AI Act becomes key. As noted by Hacker[90], socially beneficial AI applications, together with safeguards such as pseudonymisation, differential privacy and transparency measures, weigh in favour of developers in the balancing exercise. At the same time, he observes that the reasonable expectations of data subjects will often not favour AI training, since individuals do not typically expect their online data to be used for that purpose.

Yet, the assessment of reasonable expectations cannot be reduced to a purely subjective inquiry into what individual data subjects happen to anticipate. Following the entry into force of the AI Act, those expectations should also be informed by the legal framework governing AI systems placed on the Union market. The obligations imposed by the AI Act contribute to shaping what individuals may reasonably expect from lawful AI development and deployment.

Under this approach, the legitimate-interest assessment is not weakened but recalibrated. The reasonable expectations component is anchored in a more concrete benchmark, namely compliance with the requirements and obligations applicable to AI systems[91]. This framework provides a blueprint for the assessment and may therefore contribute to increasing legal certainty while supporting innovation and competitiveness.

Critics of the Digital Omnibus have voiced discontent because of the alleged weakening of data subjects, most notably of the exercise of their rights. If this critique can perhaps be supported in some

[88] Jia et al. (2021); Frey & Presidente (2024).
[89] European Law Institute (2025). Simpler, Fairer, Making Simplification and Improvement Come True – ELI's Proposed Revisions to the Digital Omnibus.
[90] Hacker, P. (2026). AI in EU Law: Training, Hallucinations, Memorization, and the Core Regulatory Architecture. SSRN, 1-25, 4.
[91] As noted by Hacker, ibid., 5, it is noteworthy that a recent judgment of the Cologne Regional Court reportedly applied the balancing test in a manner favourable to AI developers: see OLG Köln, Case 15 UKl 2/25, NJW 2025, 3156.

particular respects (for example, in respect of the abusive exercise of the right of access), this does not seem to be the case when it comes to the use of legitimate interest for processing personal data, against which individuals can "unconditionally" object.

The proposed right to object is framed as "unconditional" in the wording of Article 88c[92]. However, this characterisation diverges from the general provision under Article 21 GDPR. Under Article 21(1) GDPR, when the data subject exercises the right to object, the controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims. Accordingly, Article 21(1) GDPR does not establish an unconditional right to object. An unconditional right to object is granted to data subjects under Article 21(2)(3) GDPR, when data are processed for direct marketing purposes. Even the EDPB and EDPS have pointed out that this safeguard "goes beyond the general right to object set out by Article 21(1) GDPR"[93] (incidentally suggesting including this right in the latter provision, rather than in Article 88c). In other words, the current framing would subject the interest in the development and operation of AI systems and models to the same regime that the GDPR reserves for direct marketing activities. Being the right to object "unconditional", it would preclude any balancing exercise based on competing legitimate interests of the controller. However, while such a special rule may be justified in the context of direct marketing, the rationale is less convincing when it comes to AI development and deployment[94], where controllers may be able to demonstrate compelling legitimate grounds capable of outweighing the interests, rights and freedoms of the data subject.

From an economic standpoint, the unconditional right to object is a structural analogue of the consent fatigue documented in Section 2.1.1: the cognitive costs of opt-out mirror those of opt-in, and the behavioural evidence predicts that most users will not exercise the right, leaving it a formal safeguard with limited practical effect.

In conclusion, the proposed Article 88c would shift the GDPR at first sight away from the paradigm of a technology-neutral piece of legislation. While this feature may signal the importance of the processing of personal data in a strategic segment of the digital market, such as AI technologies, the provision contributes only marginally to legal certainty. As noted above, the balancing exercise inherent in the legitimate interest assessment is largely intact and will continue to play a crucial role for the lawfulness of the processing activities and is likely to continue generating disputes in the coming months before supervisory authorities and courts (at least in the absence of robust guidelines that may come from the cooperation between the EDPB and the European Commission on the interplay between the GDPR and the AI Act). Moreover, various aspects of the proposal leave important interpretative questions unresolved. Nevertheless, it is worth stressing that, by expressly recognising legitimate interest as a legal basis for the development and operation of AI systems and models, Article 88c reflects a clear policy preference against requiring (large-scale) consent collection

[92] The safeguards established by the proposed Article 88c in the second paragraph include, among the others, a requirement of "enhanced transparency to data subject", which however remains undefined, as noted also by the EDPB and EDPS (2026b), para. 43).
[93] EDPB-EDPS (2026b), para. 42.
[94] Another controversial aspect of the proposed provision concerns the notion of "operation" of AI systems and models. Unlike the concept of "deployment", which is expressly defined in the AI Act, the notion of "operation" is not codified anywhere in EU digital law. By departing from the terminology used in the AI Act, Article 88c raises questions as to whether "operation" and "deployment" should be regarded as overlapping concepts or whether the former encompasses a broader set of activities. The answer is likely to affect the scope of the processing activities that may rely on legitimate interest under the proposed provision. See also the observations on this issue in the reports by the European Law Institute (2025), at 18, and noyb (2026a), at 78.

for AI training purposes. Even though the proposal leaves room for EU or national legislation to require consent in specific circumstances, it confirms that such processing activities may, in principle, rely on legitimate interest.

4.2.2 Residual and incidental processing of special categories of data

A second significant proposal concerning AI model training in the Digital Omnibus relates to the regime governing the processing of special categories of personal data in the context of the development and operation of AI. The topic has already prompted academic reflections in recent months[95], particularly when the collection of personal data is based on web scraping[96].

First of all, the Digital Omnibus proposal aims to add a new point to the list of derogations from the prohibition to process special categories of personal data under Article 9(2) GDPR. The proposed exception (lit. k) would cover the processing "in the context of the development and operation of AI system […] or an AI model". At first sight, the amendment would appear as a broad exception covering every processing of special categories of personal data in both AI development and AI deployment. Only when reading the proposed amendment in conjunction with Recital 33 of the proposal does one realise that the scope of the exception is more limited, circumscribed to the residual and incidental processing of special categories of personal data "that are not necessary for the purposes of the processing" (which is not mentioned in the enacting terms of the proposal[97]). Coherently, the application of Article 9(2)(k) is made subject to the conditions referred to in paragraph 5, which would likewise be introduced ex novo by the Digital Omnibus. The provision is admittedly complex but helps clarify the scope of the derogation from the general prohibition to process special categories of data.

First, Article 9(5) requires the controller (which one can assume to be either a provider or a deployer under the AI Act) to implement appropriate organisational and technical measures to avoid the collection and other processing of special categories of personal data. This wording significantly circumscribes the scope of the new legal basis, as it makes clear that the processing of special categories of personal data should not take place by default – something that Article 9(2)(k) per se does not specify.

The second part of the provision further clarifies and refines the scope of application of the derogation, which will cover situations where, despite the implementation of the referred technical and organisational measures, "the controller identifies special categories of personal data in the datasets used for training, testing or validation or in the AI system or in AI model". In this case, the proposal provides that "the controller shall remove such data". Accordingly, the new exception does not contain a general authorisation to process special categories of personal data; on the contrary, it

[95] Kuru, T. (2024). Lawfulness of the mass processing of publicly accessible online data to train large language models. International Data Privacy Law, 14(4), 326–351.

[96] As noted by Raposo, V.L. (2026). The AI Gospel according to the EDPB–An overview of opinion 28/2024 on data protection aspects in AI models. Computer Law & Security Review, 61, 106300, 1-20, at 17, the EDPB Opinion 28/2024 does not specifically address the case where special categories of personal data are processed for AI model training. However, the EDPB reiterates the finding of the Court of Justice in the *Meta v. Bundeskartellamt* judgment that, where special categories of data are collected en bloc together with personal data that do not fall within Article 9 GDPR, the entire processing operation must rely on one of the derogations under Article 9(2).

[97] See EDPB-EDPS (2026b), para. 48.

only concerns the incidental and residual processing of special categories of data, as clarified in Recital 33.

Finally, the proposed amendment also addresses the case where the removal "requires disproportionate effort": under such circumstances, "the controller shall in any event effectively protect without undue delay such data from being used to produce outputs, from being disclosed or otherwise made available to third parties".

The multilayered structure of this provision rests on a clear starting point: the assumption that special categories of personal data can be fully excluded from AI training datasets is increasingly difficult to sustain from a technical perspective. Effective AI model development depends on large-scale and context-rich datasets, in which the complete removal of personal data at the pre-processing stage would significantly undermine model performance. While targeted filtering techniques can eliminate specific categories of data, contextual anonymisation remains subject to practical limitations. This creates a tension between the GDPR's general prohibition on processing special categories of personal data and the technical demands of effective AI model training. In addition to that, EDPB Opinion 28/2024 suggests that even the collection of marginal pieces of "sensitive" personal data together with non-special categories of personal information in training datasets would subject the entire processing operations to Article 9 GDPR[98]. Furthermore, the Court of Justice itself has offered a broad interpretation of the notion of special categories of personal data[99] while it has endorsed a strict construction of the exception relating to data "manifestly made public" by the data subject[100]. Against this background, both regulators and policy makers have progressively acknowledged the existence of "residual" or "incidental" processing of special categories of data, shifting the focus from an absolute prohibition towards the design of lifecycle-based safeguards aimed at minimising risks where preventing the collection of, or deleting, such data is not feasible. This technical constraint is reflected in the design of the amendment. While it could be contended that this provision would in turn confirm a shifting away of a technology-neutral approach traditionally underlying the GDPR, this should be nuanced at the same time in view of the ECJ case law (e.g., *Google v CNIL* case[101]) which already allowed similar exceptions to the rules on sensitive data processing in the context of online search.

In the end, the proposed paragraph 5 of Article 9 GDPR incorporates a complex set of layers: first of all, the special categories of personal data should not be processed as a default rule (which specifies and at the same time limit the scope of the proposed lit. k); and they should be in any case removed when it comes to incidental and residual processing; however, if their removal implies disproportionate efforts, the controller should limit their subsequent processing, at least in the outputs generated by AI systems.

The combination of the proposed lit. k) and paragraph 5 more precisely circumscribes the ability of controllers to process special categories of personal data in the development and operation of AI systems and models. On the one hand, these amendments reinforce the general ban on the processing of special categories of personal data, particularly given that AI training datasets are often compiled through large-scale web scraping techniques; on the other hand, echoing the reasoning of the Court

[98] Raposo (2026), at 17; see also EDPB (2024a), para. 17 and ECJ, *Meta v. Bundeskartellamt,* para. 89.
[99] See among others ECJ, Case C-184/20, *Vyriausioji tarnybinės etikos komisija,* judgment of 7 October 2022; Case C-21/23, *Lindenapotheke*, judgment of 24 January 2025; see also European Law Institute (2025), at 23.
[100] ECJ, *Meta v. Bundeskartellamt,* para. 77.
[101] Case C-136/17, *GC and Others v Commission nationale de l'informatique et des libertés (CNIL)*, judgment of 24 September 2019.

of Justice in the *Google v CNIL* case[102], they implicitly acknowledge that in complex multi-party data processing the use of special categories of personal data may be difficult to avoid – as recognised earlier in CNIL's guidance[103].

As noted in the EDPB-EDPS Joint Opinion 2/2026[104], this means that if "the processing of special categories of data is necessary for the purposes of the processing in the context of the "development and operation" of AI systems or models, the derogation will not apply and data controllers will need to rely on another derogation under Article 9(2) GDPR, if applicable. In addition, scholars have called for a further refinement of the proposed amendments with a view to limiting the derogation for the processing of special categories of personal data to the circumstances where the latter is necessary for this purpose, which should be in turn recalibrated to cover "what is strictly necessary to develop and modify AI models, not AI systems" as well as to exclude the operation of AI systems[105]. This proposal would further narrow down the scope of the exemption introduced by the Digital Omnibus. On the other hand, given the practical challenges posed by AI technologies for the enforcement of the right to erasure that will be discussed in para. 4.3., the provision interestingly reveals the awareness of the EU institutions about the possibility of residual and incidental processing, which may even consist in the memorisation of personal data.

From an economic standpoint, the multi-layered structure of Article 9(5) (*avoid–remove–protect*) has substantive distributional implications that deserve attention. The provision imposes a sequential cost on controllers: ex-ante measures to prevent collection (technical filters, dataset curation), ex-post identification of residual special categories (audit and detection capabilities), and conditional removal or protection of identified data (additional engineering and documentation). For large model developers operating at scale, these costs can be partially absorbed through dedicated privacy-engineering teams and the existing pipeline of privacy-enhancing technologies. For smaller AI developers, particularly those building on foundation models or operating with limited legal and technical resources, the layered obligation imposes compliance costs that scale less than proportionally with firm size, which is a pattern consistent with the broader empirical evidence on the GDPR's incidence reviewed in Section 2.2.1. Moreover, the "disproportionate effort" exception introduces an implicit cost-benefit test, but its decentralised application (each controller assesses what counts as disproportionate, with limited supervisory guidance) converts what appears to be a relief mechanism into a source of legal risk. The combined effect is to favour controllers with mature compliance infrastructure over those that lack it, partially offsetting the simplification gains the Digital Omnibus articulates for SMEs.

4.2.3. The Digital Omnibus on AI: a reshaped debiasing exception

Interestingly, the Digital Omnibus on AI also introduces a provision on the processing of special categories of personal data under proposed Article 4a, which refers to the specific purposes of bias detection and mitigation while defining its scope of application. This provision is expected to replace and revisit the current Article 10(5) AI Act, the wording of which is largely reproduced by the first

[102] Ibid.
[103] See CNIL, Les fiches pratiques IA. Available at: https://www.cnil.fr/fr/les-fiches-pratiques-ia.
[104] EDPB-EDPS (2026b), para. 48.
[105] Hacker (2026), at 7.

paragraph of Article 4a, and thus to autonomously regulate the processing of special categories of personal data for those purposes. As noted in the AI Act (Recital 70), Article 10(5) – and, arguably, the new Article 4a – identifies a specific case of processing for reasons of substantial public interest, one of the derogations from the general ban on processing special categories of personal data listed in Article 9(2)(g) GDPR. The most important added value of this provision lies in para. 2, which extends the possibility to rely on this legal basis beyond high-risk AI systems and therefore amends the current content of Article 10(5) AI Act. While the latter provision only authorises providers of high-risk AI systems to process special categories of personal data for purposes of bias detection and correction, the new version also empowers deployers to do so and expands the scope of the exception to AI systems other than high-risk systems.

The first paragraph of Article 4a confirms the regime and the conditions under Article 10(5) AI Act for providers of high-risk AI systems to process special categories of personal data for bias detection and mitigation. The second paragraph, however, provides that providers and deployers of AI systems and models as well as deployers of high-risk AI systems may exceptionally process special categories of personal data where two conditions are met: (i.) the processing is strictly necessary to ensure bias detection and correction in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under EU law (especially where data outputs influence inputs for future operations); and (ii.) all the conditions and safeguards set out in para. 1 for providers of high-risk AI systems are met[106].

This amendment reflects certain proposals advanced in literature to expand the scope of Article 10(5) AI Act[107] so as to ensure that also developers of low-risk AI systems are not discouraged from processing special categories of personal data for fear of violating Article 9 GDPR[108], given that biases cannot be confined to high-risk AI systems alone.

From an economic standpoint, the extension of the debiasing exception beyond high-risk AI systems is consequential in its distributional incidence. Bias detection and mitigation require access to special categories of personal data (demographic, linguistic, ethnic, contextual) that under the current Article 10(5) AI Act are accessible only to providers of high-risk systems. Restricting the exception in this way imposes a structural penalty on the very category of developers, typically smaller, often serving niche or culturally specific applications, that depend on bias-aware development to compete with larger generalist platforms. The extended scope of Article 4a addresses this asymmetry: it lowers the legal barrier to bias detection across the AI ecosystem and aligns the regulatory architecture with the empirical evidence, reviewed in Section 2.2.3, that algorithmic bias is not confined to high-risk applications[109]. The economic significance is particularly acute for the EU AI ecosystem, where adequate representation of linguistic and cultural diversity[110] requires access to demographic data that fall within Article 9 special categories. Without an extended Article 4a derogation, SME and culturally specific AI developers would face a compounded constraint, as they would be exposed to

[106] The latest version of the provision significantly departs from the Commission's original proposal. Under the latter, providers and deployers of AI systems and models other than high-risk AI systems, as well as deployers of high-risk AI systems could process special categories of personal data for the purposes of bias detection and correction where such processing was necessary and proportionate, provided that the safeguards applicable to providers of high-risk AI systems were respected.
[107] Hacker, P., Kilian, R., & Costas, J. (2025). "Simplifying" European AI Regulation. An Evidence-based Study. BartelsmannStiftung, 41.
[108] van Bekkum, M., & Zuiderveen Borgesius, F. (2023). Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?. Computer Law & Security Review, 48, 105770, 1-12.
[109] Buolamwini & Gebru (2018); Bender et al. (2021).
[110] Joshi et al. (2020).

bias risks (and to the resulting reputational and legal consequences) without legal access to the data needed to detect and mitigate those risks.

The final wording of Article 4a AI Act also responds to several observations raised by the EDPB and EDPS in their Joint Opinion 1/26[111]. First, para. 1 is now aligned with Article 10(5) insofar as the processing of special categories of personal data by providers of high-risk AI systems is permitted only where "strictly necessary" for bias detection and mitigation, whereas the European Commission's proposal authorised such processing merely "to the extent necessary" for that purpose. Second, the new version of para. 2 de facto incorporates the requirement that biases likely to affect health and safety, fundamental rights, or discrimination prohibited under EU law, thereby reflecting the logic underlying Article 10(2) (f) and (g) AI Act[112], even though those provisions are expressly referenced only in relation to providers of high-risk AI systems only[113].

Overall, Article 4a strikes an interesting balance between data protection, non-discrimination and innovation objectives. On the one hand, it broadens access to a legal basis that may be indispensable for detecting and correcting discriminatory outcomes across the AI value chain, reducing possible asymmetries that may affect certain actors. On the other hand, the adoption of a strict necessity requirement and the limitation of the derogation to situations involving risks to health, safety, fundamental rights, or discrimination ensure that the processing of special categories of personal data remains exceptional and properly circumscribed.

4.3. The enforcement of data subjects' rights under the GDPR and the challenges posed by AI technologies

4.3.1. Enforcing the right to erasure in AI models

Data subjects' rights occupy a central position in the overall architecture of data protection law, given the inherent tension between some of its core principles, such as data minimisation, accuracy and purpose limitation, and the data-intensive nature of AI systems, which strongly relies on the ability to process vast amounts of personal data at various stages. In the context of the development of AI models that require large and accurate datasets to learn "more quickly" and "better" how to identify patterns and correlations, data subjects' rights crucially empower individuals to regain control over their personal data, to the extent this is possible. Recent, albeit limited, developments in this domain have largely been driven by the European Court of Justice, which released two landmark judgments concerning the right not to be subject to automated decision-making under Article 22 GDPR in the *Schufa*[114] and *Dun & Bradstreet*[115] cases. These judgments emphasised the need to adopt a broad interpretation of Article 22 GDPR to safeguard individuals against "too much automation" and the

[111] EDPB-EDPS (2026a). Joint Opinion 1/2026 on the Proposal for a Regulation as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI), para. 12.
[112] Ibid, para. 13.
[113] The latter reference implies that the processing in question occurs to comply with specific obligations currently imposed on providers of high-risk AI systems, i.e., to adopt governance and management practices in relation to testing and validation data sets, and most notably the "examination in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law" and "appropriate measures to detect, prevent and mitigate possible biases identified" accordingly.
[114] ECJ, Case C-634/21, *OQ v Land Hessen,* judgment of 7 December 2023.
[115] ECJ, Case C-203/22 - Dun & Bradstreet Austria, judgment of 27 February 2025.

inherent connection between the data subjects' right of access and the contestability of the output of automated decision-making[116].

The Digital Omnibus addresses some practical aspects, ranging from the abusive use of the right of access to some exemptions from the duty to inform data subjects. However, it does not delve into a question of huge practical impact for AI providers, namely the enforcement of the right to erasure, on which an increasing demand for legal certainty has arisen[117]. The EDPB already tackled, albeit incidentally, some of these problems[118] in its Opinion 28/2024[119], following interventions by national DPAs[120].

A pivotal and preliminary issue in the application of the GDPR to AI concerns whether trained models themselves can be regarded as anonymous and therefore fall outside the scope of application of data protection law. This point is key to understanding whether data subjects' rights can be enforced against the controller/model developer. According to Opinion 28/2024, AI models trained on personal data cannot, in all cases, be considered anonymous. This position rejects a line of argument occasionally advanced in technical and legal literature[121], according to which trained models merely encode statistical relationships and therefore do not "contain" (and therefore do not process) personal data. Instead, the EDPB adopts a functional and risk-based approach, aligned with the broad notion of personal data under the GDPR. Model anonymity is not impossible; however, it cannot be presumed and must be demonstrated and assessed on a case-by-case basis. To this end, Opinion 28/2024 articulates a two-pronged test, based on which an AI model may qualify as anonymous only where it is very unlikely that a) personal data can be extracted – directly or indirectly – from the model (including through techniques such as model inversion or membership inference); b) personal data can be obtained through interaction with the model, even unintentionally (e.g., via prompts or queries). Both these conditions must be satisfied.

Interestingly, the test goes beyond the internal model structure to include its external behaviour (such as the output it may generate when prompted). Additionally, Opinion 28/2024 clarifies that both the likelihood of extraction of personal data and that of obtaining personal data from a model must be assessed taking into account "all the means reasonably likely to be used" by the controller or another person and should also consider unintended (re)use or disclosure of the model. This formulation, derived from Recital 26 GDPR, has several implications in the AI context. Indeed, it does not only require considering state-of-the-art attack techniques (e.g., extraction attacks), but also possible actions by third parties beyond the controller and its power of control. It also requires a situational and context-based assessment that may evolve and change over time.

[116] Binns, C., & Veale, M. (2021). Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR. International Data Privacy Law, 11(4), 319-332, who emphasise this tension especially in the context of multi-party processing.
[117] Vrabec, H. (2025). Erasing personal data in an AI era. Leidenlawblog.
[118] See also EDPB - 2025 Coordinated Enforcement Action (2026). Implementation of the right to erasure by controllers. Adopted on 10 February.
[119] Raposo (2026). More recently see also Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Data (Convention 108) (2026). Draft Guidelines on Privacy and Data Protection in the context of LLM-based systems. Council of Europe, T_PD(2025)3rev3, 12 May.
[120] Datatilsynet (2023). Offentlige myndigheders brug af kunstig intelligens Inden I går i gang. 1-39; Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit (2024). Large Language Models und personenbezogene Daten. 1-10.
[121] Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit (2024).

The key point made by the EDPB lies in the conclusion that model anonymity cannot be assessed in purely abstract terms[122]. This conclusion depends on both the fact that a model accessible to the public may present higher risks than a purely internal model and the circumstance that the very same model could be considered anonymous or not, based on the deployment context and environment. In the EDPB's view, it is up to controllers to demonstrate that the risk of identification is immaterial, including by resorting to appropriate documentation and the implementation of adequate safeguards, such as measures to prevent memorisation in the model design and the adoption of best practices in data collection.

In light of this interpretation, model anonymity is likely to constitute the exception rather than the rule, with the consequence that AI models will often remain within the scope of application of the GDPR, regardless of whether model training entails the actual storage of the content of training datasets. Against this background, the enforcement of the right to erasure under Article 17 GDPR[123] – commonly referred to as the "right to be forgotten" –therefore becomes particularly complex when applied to AI systems and models. While the provision is relatively straightforward in traditional data processing operations, its application in the context of AI technologies brings to light a structural tension between the logic of data subject rights (i.e., preserving individuals' control over the processing of their personal data) and the technical architecture of machine learning. In its traditional formulation, the right to erasure reflects the idea that individuals should retain control over the temporal dimension of data processing, so that personal data are not retained indefinitely once they have served the original purpose of their collection. However, this paradigm encounters significant difficulties when transposed to AI applications, because of the peculiar way in which data are processed and may be memorised. Machine learning techniques typically rely on large datasets used during the training phase to produce models that encode statistical relationships rather than storing data in a directly accessible form. Even where the original training data were deleted, the model may retain representations of those data, and in some cases may reproduce or even "memorise" fragments of the training datasets. The problem is amplified by the fact that the influence of individual data points is distributed across the model's parameters, making it technically difficult to isolate and remove the contribution of a specific piece of information. Furthermore, there is uncertainty on the legal status of inferences from a data protection perspective[124].

These features challenge the well-established understanding of erasure and undermine the potential of this data subject right in preserving individuals' control. In a conventional database, deletion entails the removal of a discrete record. In a trained model, by contrast, deletion is not a simple operation, as data have been transformed into weights and relationships that are neither directly identifiable nor easily reversible. As a result, compliance with Article 17 cannot be satisfied merely by deleting the original information from the training datasets. The more difficult question is whether, and to what extent, the right to erasure requires a technical intervention at the level of the model itself.

[122] Holzenberg, N. & Maxwell, W. (2025). A quantitative approach to the GDPR's anonymisation and ''appropriate technical and organisational measures'' tests. Computer Law & Security Review, 59, 106173, 1-13.
[123] Ausloos, J. (2020). The Right to Erasure in EU Data Protection Law. Oxford University Press.
[124] Wachter, S., & Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Columbia Business Law Review, 2, 494-620.

This question has led to increasing attention to the concept of "machine unlearning", understood as a set of techniques aimed at removing the influence of specific data from a trained model[125]. From a legal perspective, a key question is whether unlearning can be seen as a functional analogue to erasure, insofar as it seeks to ensure that personal data are no longer present in, or recoverable from, the system. Yet, the current state of the art reveals important limitations in this technique[126]. If unlearning is understood as the complete removal of any influence that a given set of personal data may have on the operation of AI systems (as in the case of "exact" machine unlearning), erasure remains technically challenging: ascertaining that data have been fully removed is difficult and, in some cases, compliance may require retraining the model from scratch, raising concerns about costs and feasibility.

For AI developers, this requires understanding what erasure means technically. There is a fundamental distinction between *removal* methods – which target the elimination of specific training data from the dataset before or during training – and *suppression* methods – which target the model's outputs, either by modifying the trained model (e.g., through additional fine-tuning or model editing) or by adopting system-level filters that prevent certain content from reaching the user[127]. These are not interchangeable approaches. *Removal* operates on observed information (i.e., training data) while *suppression* operates on outputs. Bearing this difference in mind, removal of specific training data does not guarantee output suppression: a model retrained without a particular individual's data may still be capable of generating outputs that reflect information about that individual[128].

This observation supports a conclusion: if neither removal nor suppression can reliably and systematically deliver the outcome that Article 17 contemplates – the elimination of an individual's data and its effects from a system – then compliance cannot be structured primarily as an ex-post remedy, but must be built into the design of the model from the outset. Retraining a model should only be considered under very exceptional circumstances, as a very last resort, i.e., in accordance with the principle of proportionality and based on technical feasibility. This approach is also reflected in guidance issued at Member State level, including by the CNIL[129]. In accordance with this approach, removal or suppression remain the preferred routes – an approach which by the way is also taken in the proposed new Article 9 (5) GDPR in the Digital Omnibus, which requires eliminating special categories of personal data subject to incidental and residual processing and, where removal would involve disproportionate efforts, to prevent them from being used to produce outputs , be disclosed or made otherwise available.

From an economic standpoint, the structural mismatch between Article 17 GDPR and AI systems generates a distorted distributional pattern. The compliance cost gradient across the AI lifecycle is sharply non-linear. For example, removing a single training record from a dataset prior to training is essentially costless — a one-line database operation. The same record removal after a foundation model has been trained may require partial or complete retraining of the model, at a cost that for

[125] Juliussen, B.A., Rui, J.P., & Johansen, D. (2023). Algorithms that forget: Machine unlearning and the right to erasure. Computer Law & Security Review, 51, 105885, 1-12.
[126] Zhang, D. et al. (2025). Right to be forgotten in the Era of large language models: implications, challenges, and solutions. AI and Ethics, 5, 2445-2454.
[127] Cooper, F.A. et al. (2025). Machine Unlearning Doesn't Do What You Think: Lessons for Generative AI Policy and Research. arXiv:2312:06966v2.
[128] Ibid.
[129] CNIL (2026). Ensuring and facilitating the exercise of data subjects' rights, 5 January. Available at https://www.cnil.fr/en/ensuring-and-facilitating-exercise-data-subjects-rights.

state-of-the-art systems can run into the order of millions of dollars in compute alone[130]. At the deployment stage, output filtering and incremental fine-tuning can be implemented at intermediate marginal cost, but as Cooper et al.[131] emphasise, these techniques deliver suppression rather than removal: the underlying data, and the latent representations they have produced during training, remain in the model. Total compliance cost is therefore minimised by ex-ante data governance rather than ex-post individual rights — a conclusion consistent with the broader argument advanced in Section 2.2 that AI risk mitigation depends on design obligations whose preconditions can be systematically met, not on individual safeguards whose preconditions are systematically unmet.

The cost gradient also interacts asymmetrically with the multi-actor structure of the AI lifecycle. Large foundation-model providers can internalise privacy-by-design infrastructure and amortise the fixed costs across users; smaller AI developers, including those building on pre-trained foundation models, often lack the technical capacity to implement removal at the model layer and depend on suppression-based workarounds whose compliance status remains unsettled. This asymmetry compounds the broader fixed-cost incidence on SMEs documented in Section 2.2.1. The implicit shift toward output filtering — which the European Commission endorses elsewhere in the Digital Omnibus through the Article 9(5) "disproportionate effort" provision – reflects a pragmatic compromise that reduces compliance cost at the price of leaving the underlying data in the model. As Cooper et al.[132] emphasise, output filtering is suppression rather than removal, and residual exposure to adversarial extraction remains. From a regulatory-economics standpoint, the broader question is whether enforcement targets observable suppressive measures or technically demanding removal – a choice that determines, in practice, the level of substantive privacy protection the regime will deliver.

These constraints indicate that, in the context of AI technologies, the right to erasure cannot be fully operationalised ex-post, but requires ex-ante design choices that minimise the incorporation and persistence of personal data. The inherent complexity of erasure is further amplified when considered across the different stages of the AI lifecycle. During the data collection phase, individuals may request the deletion of data used for training. At the deployment stage, erasure may concern outputs generated by the system or logs associated with its operation. These layers of processing often involve multiple actors, including providers and deployers, whose responsibilities may overlap or diverge. This fragmentation complicates the allocation of obligations and, ultimately, the enforcement of data subjects' rights at large. In addition, the exercise of the right to erasure must be balanced against competing interests, notably freedom of expression and information[133]. This tension is particularly acute when it comes to AI systems used to generate or retrieve content derived from publicly available sources, where erasure requests may conflict with the societal interest in access to information, requiring a careful balancing exercise, particularly when it comes to public figures. Against this background, output filtering (which the European Commission's proposal suggests applying in case of incidental and residual processing of special categories of personal data, where their deletion would require disproportionate efforts) should be carefully evaluated as a solution considering the relevant context-based impact.

[130] Cooper et al. (2025).
[131] Ibid.
[132] Cooper et al. (2025).
[133] noyb (2024). ChatGPT provides false information about people, and OpenAI can't correct it.

Taken together, these elements describe a critical intersection between the GDPR and AI systems, which the forthcoming guidelines from the European Commission, the EDPB and the EDPS on the interplay between the two regulations should address. The right to erasure (and, similarly, also the right to rectification) presupposes a model of data processing based on identifiable data points and reversible operations. AI models, by contrast, rely on distributed representations and iterative learning. This does not render the right to erasure inapplicable, but it does challenge its operationalisation. Rather than operating solely as an individual corrective mechanism, it increasingly acts as a systemic constraint, encouraging controllers to adopt privacy-preserving design strategies, limit the use of personal data and develop technical solutions that enable compliance (such as output filtering, where appropriate)[134]. However, as noted above, output filtering constitutes suppression rather than removal in the technical sense, and policy makers and supervisory authorities should calibrate their expectations accordingly. Filtering can reduce the risk that personal data are visible to users, but cannot ensure that the underlying data have been deleted from the model or that they could not be elicited through adversarial means. In this sense, the right to erasure in the AI context should be understood not merely as an ex-post remedy, but as a prerequisite shaping the design and governance of AI models.

In the end, the technical obstacles to effectively enforce the right to erasure in the context of AI models and systems may lead to two conclusions. First, more clarity and guidance are needed in this specific respect. If the impossibility to effectively operationalise the erasure from a technical perspective cannot be an excuse for exempting data controllers from their obligations under the GDPR, the currently available guidance provides limited certainty for AI operators and may generate compliance costs that are difficult to anticipate ex ante, potentially undermining innovation. The ECJ's 2014 ruling on search engine providers offers a relevant precedent in this respect. By accepting the de-indexing of specific URLs as a legally sufficient form of erasure, the Court demonstrated that technology-neutral provisions can be interpreted in light of the technical architecture of the system at issue – effectively treating what is, in technical terms, an instance of output suppression. A similar interpretive exercise will be required as supervisory authorities and courts are called upon to determine what Article 17 demands in the context of trained AI models. A second point concerns the way the right to erasure can be operationalised: if it turns out that data removal is technically impossible to fulfil, a shift from ex-post intervention to ex-ante data governance becomes necessary. In other terms, the difficulties in enforcing this right should demand a more careful management of personal data in the training phase. In this respect, Article 10 AI Act, which establishes a data governance and data quality requirements for high-risk AI systems[135], may play an important role in shaping the processing of personal data during the training phase[136].

4.3.2. The right of access as a contestation-enabling safeguard?

One of the aspects that has raised the most criticism in the discussions following the European Commission's proposal on the Digital Omnibus concerns the limitations in the scope of application of the right of access that are mediated through the restrictions on the information rights under Article 12 GDPR. As is well known, this provision defines the modalities for the exercise of the data subject

[134] Cooper et al. (2025).
[135] Hacker, P. (2021). A legal framework for AI training data – from first principles to the Artificial Intelligence Act. Law, Innovation and Technology, 13(2), 257-301.
[136] Hacker et al. (2025).

rights. By doing so, in its current wording, para. 5 provides that both the provision of information under Articles 13 and 14 and any communication relating to the exercise of the rights under Articles 15-22 shall be free of charge. It only carves out an exception for cases of requests that are manifestly unfounded or excessive (in particular because of their repetitive character): under such circumstances, the controller may either charge a reasonable fee or decline the request. In any case, the controller bears the burden of proof of the manifestly unfounded or excessive character of the request.

The European Commission's proposal on the Digital Omnibus aims to extend the scope of the exception to the circumstances where, for requests under Article 15, "the data subject abuses the rights conferred by this Regulation for purposes other than the protection of their data". This suggested amendment has received heavy criticism from scholars[137], civil society and consumer organisations[138] alike, inter alia because of its expected impact on the contestation-enabling nature of the right of access that was recently emphasised by the ECJ in the *Dun & Bradstreet* judgment, most notably in the age of AI[139]. The connection between the AI domain and the role of the right of access is already visible in the GDPR, to the extent that, on the one hand, Articles 13(2)(f) and 14(2)(g) and, on the other hand, Article 15(1)(f) include among the obligations for controllers (and, thus, within the data subject right) the provision of information on the existence of automated decision-making (extending to meaningful information about the logic involved, the significance and the envisaged consequences for data subjects).

Scholars have characterised the right of access as an emancipatory tool, the purposes of which go beyond "the idea of access as a mere tool that allows people to verify the lawfulness of the processing of personal data"[140]. As a matter of fact, requests under Article 15 are frequently relied upon also to facilitate the enforcement of other rights, for example by journalists and activists[141], or to empower vulnerable groups such as workers in accessing and contesting existing data practices[142]. Given the lack of more structured individual remedies in the AI Act[143], it is likely that society at large will develop significant expectations regarding the potential of the right of access for broader scrutiny and oversight of AI-driven practices and applications. Additionally, NGOs have advanced the claim that empirically most access requests are not properly answered by controllers[144].

The proposed amendment to the wording of Article 12(5) is therefore claimed to weaken not only the fundamental right to data protection but also the avenues it offers to individuals and groups to pursue other societal interests. This would be in contradiction with the GDPR's declared purpose to protect not only natural persons with regard to the processing of their personal data but more generally their fundamental rights and freedoms[145]. The critique of the amendment to Article 12(5) therefore has

[137] Aloisi A. (2026). "Undo something. Whatever it takes". The impact of the EU Digital Omnibus on workers. SSRN Working Paper, 1-28; González Fuster G. (2026); Mahieu, R. (2025). The Ominous Omnibus: Dismantling the Right of Access to Personal Data. Verfassungsblog.
[138] noyb (2026a); BEUC (The European Consumer Organisation) (2026). Response to consultation Protecting EU data and privacy rights in the Digital Omnibus.
[139] Metikoš L. (2025). Dun & Bradstreet: A Pyrrhic Victory for the Contestation of AI under the GDPR. KU Leuven blogpost; Rossetti, S. (2025). The Court of Justice of the European Union Confirms the Existence of the Right to Explanation of Automated Decision-Making. European Law Blog.
[140] Mahieu, R. (2023). The right of access to personal data in the EU: a legal and empirical analysis. Doctoral thesis. Vrije Universiteit Brussel, at 318.
[141] Mahieu (2025).
[142] Aloisi (2026).
[143] Paolucci F. (2024). Shortcomings of the AI Act. Evaluating the New Standards to Ensure the Effective Protection of Fundamental Rights. Verfassungsblog.
[144] noyb (2026b).
[145] EDPB-EDPS (2026b), para. 54.

reasonable grounds as massive refusal practices adopted by controllers may have potentially systemic effects in an AI-driven society. However, the gap between the actual purposes pursued by the amendment and the unintended consequences it is claimed to produce is narrower than one could imagine.

A closer look at the proposal, however, suggests that some of these concerns may be overstated. Indeed, looking at Recital 35 of the Digital Omnibus proposal, one can notice that the objective is not that of weakening the potential for contestation inherent in the right of access. The recital provides that the right of access "should not be abused in the sense that the data subjects abuse them for purposes other than the protection of their data". Taken alone, this wording might prima facie suggest that an abuse takes place every time the right is exercised for purposes other than data protection. One might therefore argue that a request of access for the exercise of freedom of information, for instance, should constitute per se an abuse[146]. However, Recital 35 further clarifies the point through the use of examples: "such an abuse of the right of access would arise where the data subject intends to cause the controller to refuse an access request, in order to subsequently demand the payment of compensation, potentially under the threat of bringing a claim for damages. Other examples of abuse include situations where data subjects make excessive use of the right of access with the only intent of causing damage or harm to the controller or when an individual makes a request, but at the same time offers to withdraw it in return for some form of benefit from the controller". These examples illustrate what kind of purposes other than data protection may be interpreted as abusive.

The goal of the Digital Omnibus should not be to place constraints on the enforcement of various fundamental rights and, for instance, weaken the grounds for strategic litigation in the collective interest. On the contrary, by allowing controllers to refuse abusive requests for access, it seems that the European Commission aims to avoid a "weaponisation" of data protection when the actual purpose does not lie in protecting personal data or other fundamental rights but rather raising instrumental claims depending on the way controllers take access requests into account.

Ultimately, neither the wording of the amendment nor the examples provided in Recital 35 necessarily suggest an intention to deprive data subjects of the potential inherent in the right of access for the protection of personal data and other fundamental rights. It is legitimate to question whether the enforcement of these rights genuinely requires manifestly unfounded or excessive access requests. Recital 35 also suggests that "[O]verly broad and undifferentiated requests should also be regarded as excessive", requiring that data subjects be "as specific as possible". In any case, controllers bear the burden of demonstrating the manifestly unfounded or excessive nature of each access request.

In the end, the added value of this modification could be quite limited, as Article 12(5) GDPR already provides for an exception in the case of abuse[147]. If, however, the goal underlying the Digital Omnibus is to reinforce the controllers' power of refusal, the current wording of the amendment (as well as Recital 35) should be refined to clarify that the refusal of access should not be interpreted in a way that prevents the enforcement of other fundamental rights, so as to preserve the "contestation-enabling" nature of the right in question particularly in sensitive AI-driven contexts. This is exactly the problem identified by the EDPB-EDPS in their Joint Opinion, which has suggested linking the notion of "abusive requests" to the existence of an abusive intention, such as that of causing harm to the

[146] noyb (2026a), at 29.
[147] Ibid., at 33.

controller[148]. This seems in line with the very recent judgment of the Court of Justice in the *Brillen Rottler* case[149]. The Court noted that a first request for access may, in certain circumstances, already be regarded as "excessive" within the meaning of the GDPR and may therefore be abusive. This may be the case when the controller demonstrates that the request was made not for the purpose of being aware of the processing of the data and assessing its lawfulness, but with the "abusive" intention of artificially creating the conditions for obtaining an advantage under the GDPR in the form of compensation (§ 45). The fact that, according to publicly available information, the data subject made a large number of requests for access to their personal data, followed by claims for compensation to various controllers, may be taken into consideration for the purpose of establishing the existence of an abusive intention (§ 45). The reasoning of the Court signals that the assessment of the abusive intention is largely context-based and situational.

At the same time, consistently with a "stick-and-carrot" approach, the European Commission should take seriously the alleged problem of enforcement of the right of access, which is reported to be not fully satisfied by controllers. Finding a proper way to safeguard data subjects while not disproportionately adding red tape for controllers seems to be a challenge ahead, which is perhaps not limited to the data protection domain in the EU digital rulebook.

From an economic standpoint, the right of access serves a function whose value extends beyond the individual data subject who exercises it. In settings of high information asymmetry — and AI systems are paradigmatic — the right operates as a decentralised mechanism for the production of accountability information that benefits not only the requester but society at large. Its public-good character is particularly substantial in the AI context, where the absence of strong alternative individual remedies[150] leaves the right of access as one of the few effective tools for ex-post scrutiny of automated decision-making. The behavioural dynamics of access requests, however, are sharply asymmetric. Consistent with the broader argument advanced in Section 2.1, the fixed cognitive and time costs of formulating a substantive access request mean that most data subjects will not exercise the right at all. Those who do are self-selected for sophistication, public-interest motivation, or strategic-litigation capacity – and even they face systematic non-compliance, with empirical evidence indicating that 83.5% of access requests are not properly answered by controllers[151]. Against this background, the proposed amendment to Article 12(5) produces an unfavourable cost-benefit configuration. The marginal benefit is modest: Article 12(5) already permits refusal of manifestly unfounded or excessive requests, and recent case-law (most recently *Brillen Rottler*) already recognises abusive intention as a ground for refusal. The marginal cost, by contrast, falls precisely on the population whose exercise of the right generates the most public-good value: investigative journalists, advocacy organisations, and individuals contesting AI-driven decisions. The expected aggregate effect is therefore a chilling effect on socially valuable uses of the right, in exchange for a marginal reduction in abuse that the existing framework already addresses through the abusive-intention test.

In addition, given the context-based nature of the assessment of an abusive intention emphasised in *Brillen Rottler*, the EDPB and EDPS are all the more right in suggesting removing the presumption that

[148] EDPB-EDPS (2026b), paras. 54-55.
[149] ECJ, Case C-526/24, *Brillen Rottler,* judgment of 19 March 2026.
[150] Paolucci (2024).
[151] noyb (2026b).

"overly broad and undifferentiated requests should be regarded as excessive"[152]: not because this is necessarily inaccurate, but rather because the specification, which may pose a risk of "over-refusal" when applied by controllers, is perhaps unnecessary.

4.4. Cookie fatigue: consent management and privacy signals

The Digital Omnibus also responds to the debates that have arisen among scholars, policy makers, and the private sector concerning so-called "cookie fatigue". There is broad consensus that EU cookie law does not effectively safeguard users' privacy, while imposing significant costs and burdens on website operators[153]. The management of data subjects' consent through cookie banners has proven both ineffective and too complex; ultimately, it fails to deliver on the promise of more informed user decisions. In response to this "cookie fatigue", the European Commission has proposed introducing two new provisions into the GDPR, namely Articles 88a and 88b. Although these provisions are not AI-specific, they operate at a layer upstream of most AI systems, namely access to and control over user data. As such, they are likely to affect data collection practices in the near future. At the same time, they provide an opportunity to reflect further on the role of consent in an AI-driven digital economy, complementing the analysis of the legal bases for the processing of personal data in the development and deployment of AI systems. In a nutshell, the key amendments, on the one hand, indicate consent as the legal basis for the processing of personal data for the access to and use of terminal equipment; on the other one, they aim to introduce a centralised consent management mechanism.

The proposed Article 88a GDPR governs the conditions under which personal data may be stored or accessed on the terminal equipment of natural persons. It identifies consent as the default mechanism for such processing under the GDPR, while allowing for limited derogations where processing is necessary for a series of purposes. Article 88a should be read in conjunction with the proposed Article 88b, which regulates automated signals. By enabling data subjects to give consent, refuse consent requests, or exercise their right to object through automated and machine-readable means, it seeks to alleviate consent fatigue in practice. Although the wording of the two provisions does not explicitly establish a connection, they are functionally intertwined and pursue the same goal.

The two provisions, however, have received quite markedly different reactions. Article 88a has been cautiously welcomed, while Article 88b has attracted substantial criticism from the industry and some Member States (e.g., Italy)[154]. More than other provisions, while these amendments appear to be motivated by the aim for simplification, they are unlikely to deliver actual benefits for data subjects and controllers.

4.4.1. A fragmented consent regime

Article 88a brings under the umbrella of the GDPR (personal) data processing activities that so far have been governed by Article 5(3) of the ePrivacy Directive. This identifies as the legal basis applicable to

[152] EDPB-EDPS (2026b), para. 56.
[153] For some background on EU cookie law, see Tomisek, J. (2023). Cookies and EU Law: History, Future Regulation and Critique. Technology and Regulation 35-44.
[154] See Camera dei Deputati - XIV Commissione (Politiche dell'Unione europea), Documento approvato dalla XIV Commissione nell'ambito della verifica di sussidiarietà di cui all'articolo 6 del Protocollo n. 2 allegato al Trattato di Lisbona, 24 February 2026.

personal data, of already stored, terminal equipment of a natural person. However, Article 88a(3) also establishes some exceptions to consent, two of which almost verbatim reproduce those provided for in Article 5(3) of the ePrivacy Directive, which reflect the assumption that consent is inherently implicit[155] in the relevant processing activities (e.g., (a) carrying out the transmission of an electronic communication over an electronic communication network; (b) providing a service explicitly requested by the data subject). The other exceptions under points (c) and (d), concerning audience measurement and security purposes, are still narrowly framed. Particularly, Article 88a(3)(c) refers to "creating aggregated information about the usage of an online service to measure the audience of such a service, where it is carried out by the controller of that online service solely for its own use", excluding any audience measurement done for third parties, whereas Article 88a(3)(d) covers the "maintaining or restoring the security of a service provided by the controller and requested by the data subject or the terminal equipment used for the provision of such service", excluding fraud detection and prevention. As noted, Article 88a(3)(d) must be contextualised within the tension between "the increasing need for cybersecurity measures" and "the historically strong protection afforded to terminal equipment as part of the user's private sphere"[156]. The exception is limited in scope to device-level security[157] and should not transform cybersecurity considerations into a general "backdoor" for the processing of personal data for other purposes such as tracking users[158].
Two main concerns have been raised so far concerning the consent mechanism and the relevant exceptions, as framed under the Digital Omnibus proposal[159].

First of all, if consent still plays a pivotal role in this domain, the list of exceptions seems inherently flawed and inconsistent to the extent other low-risk processing activities have not been considered, and therefore the new proposal risks to become very soon outdated – with an article that remains too prescriptive at the basis. Indeed, the fact that various low-risk activities are left outside the Article 88a(3) whitelist looks difficult to reconcile with a risk-based approach that should be a guiding blueprint also for the Digital Omnibus. This is not merely a matter of privacy but rather (and mostly) of competition[160]. Activities such as frequency capping, fraud detection and prevention, traffic validation, and functional personalisation based on first-party data all involve limited privacy risks, which in many cases are materially lower than those arising from the data aggregation practices that the proposed exemptions aim to accommodate[161]. The EDPB and EDPS have themselves recognised the case for an additional exemption for contextual advertising[162]. Similarly, the list of exceptions could be further enriched on the basis of national data protection authorities' decisions to exempt some use cases from the consent requirement. Overall, the European Commission's decision not to

[155] noyb (2026a), at 65.
[156] Aldibs et al. (2026), at 39.
[157] Ibid.
[158] See noyb (2026a), at 70, suggesting that merely theoretical security purposes may be invoked to justify intrusions into potentially sensitive information. See also Pronesti, D. (2026). Device-Access under the Digital Omnibus: More Security or Just Less Privacy?. KU Leuven CiTiP blog, 24 March. Interestingly, the EDPB and the EDPS (EDPB-EDPS (2026b), para. 103) have called for a better specification of the exception to clarify that it extends to IT security and data protection security. Joint Opinion 2/2026 supports this exception for the purpose of ensuring the security of a service or terminal equipment, although it indicates some conditions for it to be validly invoked as a legitimate interest derogating from consent.
[159] Höppner, T., and Werle Y. (2026). Digital Omnibus: What Would it Mean for Competition and Privacy in Advertising?. The Platforms Law Blog, 9 March.
[160] Ibid.
[161] Calls to introduce other exemptions in the whitelist have been advanced by various industry associations but also in academic literature: see Höppner & Werle (2026); see also Ecommerce Europe (2026). Ecommerce Europe Position Paper on the Digital Omnibus, 23 March; IAB Europe (2026). IAB Europe's Position on the Draft Digital Omnibus on the Digital Acquis, 23 February; Alliance Digitale (2026). European Commission's Digital Omnibus (Digital Package on Simplification) Position paper, March.
[162] EDPB-EDPS (2026b), para. 103.

include such low-privacy activities is difficult to reconcile with the proposal, in Article 88c, to identify a legitimate interest in the processing of personal data for the development and operation of AI systems – a category of processing that, comparatively speaking, carries greater risks than frequency capping or fraud detection. If the Digital Omnibus, through more simplification, ultimately aims to boost European competitiveness, a limited set of exceptions to consent in this domain does not fully align with this goal.

Second, a significant inconsistency arises from the only partial integration of Article 5(3) of the ePrivacy Directive into the GDPR. As Article 88a covers only personal data processing, accessing non-personal data on terminal equipment should remain governed by the ePrivacy Directive. This creates an outcome described as paradoxical[163]: access to non-personal data on a device falls under the ePrivacy framework's stricter consent regime, while access to personal data is governed by the GDPR's more flexible (and future-proof) framework[164]. As the EDPB and EDPS noted in their Joint Opinion 2/26,[165] this partial integration risks generating legal uncertainty for controllers who would need to determine, prior to each data access, whether the data qualifies as personal. A shift from the consent-strict necessity rationale behind the ePrivacy Directive and a full alignment to the GDPR and its more flexible approach would therefore be beneficial and desirable to provide greater clarity and legal certainty.

4.4.2. The six-month ban: a paradoxical option?

The proposed Article 88a(4) defines the safeguards applicable when the legal basis for the processing of data is consent. First of all, in such circumstances, the data subject shall be able to refuse a request for consent in an easy and intelligible manner with a single-click button or equivalent means. Secondly, once they have given their consent, the controller shall not make a new request for consent for the same purpose for the period during which the controller can lawfully rely on such consent. Additionally, if a request for consent is declined, the controller shall not make a new request for the same purpose for a period of at least six months. The same conditions apply to the subsequent processing of personal data based on consent.

The six-month ban on renewed consent requests raises concerns of consistency with the rationales behind EU data protection law, most notably where combined with technical solutions such as a single-click button. This way, the provision fails to account for legitimate changes in the data processing context – such as modifications to purposes, new business partners, or significant changes to the services offered – that would normally warrant a new consent request. A rigid lock-in period is difficult to reconcile with the GDPR's risk-based approach as well as the inherent business dynamism in a data-driven and data-intensive economy. As noted, requiring an informed consent seems to conflict with prohibiting – or in any case limiting – the very request through which it is obtained[166]. The mechanism only prima facie accommodates users and alleviates their cookie fatigue; and in any case, even if it (partially) contributes to this end, it does so by significantly weakening the individuals' control over data that is generally associated with consent, ultimately limiting their autonomy – a point that will be further elaborated upon in the next section on the proposed Article 88b as well.

[163] Höppner & Werle (2026).
[164] Ibid.
[165] EDPB-EDPS (2026b), para. 97.
[166] Höppner & Werle (2026).

In their Joint Opinion, the EDPB and EDPS have recommended including a maximum period of validity of consent[167], which could be aligned with the six-month cooling off period. The suggestion touches upon a widely debated issue, namely whether validly given consents expire at a certain time[168]. The GDPR obviously subjects the validity of consent to some conditions (e.g., by empowering data subjects to withdraw their consent or through the principle of purpose limitation), however the point of whether consent should be renewed after a certain period (which implies its expiration) has been discussed by data protection authorities. From a general point of view, a statutory definition of the period of validity for consent in this domain could provide more legal certainty for both data subjects and controllers. The EDPB and EDPS themselves indicate that this provision could remind data subjects of their consent choices at appropriate intervals. However, insofar as data subjects are entitled to withdraw consent, introducing an "expiration date" does not seem either necessary or entirely consistent. Additionally, if taken together with the consent ban, an expiration date (most notably if identified in six months) would further weaken the ability of controllers to rely on consent and facilitate data subjects' refusals without actually raising their level of awareness.

Besides these theoretical inconsistencies, concerns have arisen regarding a possible informational paradox, where to record and comply with a refusal to consent, controllers may end up processing more data than in ordinary circumstances where consent is refused. Such an effect would determine a visual conflict with the principle of data minimisation. To this end, the EDPB and EDPS[169] have recommended that the recording of the refusal of consent should involve the use of generic information, such as a flag or code, common to all data subjects who have refused consent[170]. In addition, industry associations have voiced that the enforcement of the cooling-off period across devices, browsers and different environments may prove very challenging from a technical perspective[171].

Like other measures, the six-month ban has an impact beyond the privacy and data protection domain, which extends to competition. As a matter of fact, enabling individuals to express their refusal through a single-click button and preventing new requests for consent for at least six months does not contribute to value the differences that may emerge in the relevant services. Such a mechanism will also affect the implementation of emerging pay-or-consent schemes, as users will be deprived of the opportunity to consent once they have opted for rejecting all the processing activities[172]. Overall, this mechanism would reduce the ability of online advertising to fund some business models and consolidate existing market asymmetries.

4.4.3. A centralised consent management mechanism

The proposed Article 88b GDPR introduced by the Digital Omnibus complements Article 88a by explicitly addressing automated and machine-readable signals. Based on the proposed Article 88b(1), controllers must ensure that their online interfaces allow data subjects to both give consent and decline a request for consent or exercise the right to object through automated and machine-readable means. Controllers that are media service providers are expressly excluded from the scope of

[167] EDPB-EDPS (2026b), para. 105.
[168] As is well known, the GDPR requires consent to be a freely given, specific, informed and unambiguous indication of the data subject's wishes but does not specify the period for which it will be valid.
[169] EDPB-EDPS (2026b), para. 105.
[170] While noyb (2026a), at 70, has emphasised that refusals should be stored anonymously.
[171] Ecommerce Europe (2026), at 4.
[172] Höppner & Werle (2026).

application. Providers of web browsers that are not SMEs are required to provide the relevant technical means to allow data subjects to give their consent and to refuse a request for consent and exercise the right to object. The European Commission is to request standardisation organisations to develop harmonised standards for the interpretation of machine-readable signals, and controllers must respect choices made through such signals.

The EDPB and EDPS[173] have expressed support for this mechanism, viewing it as the most effective available tool for alleviating cookie fatigue by enabling users to express their preferences once across all the visited services. They have recommended clarifying the binding nature of automated signals on all controllers – not only those operating online interfaces but also third-party cookie providers – extending the obligation to web browser providers that are SMEs (as well as to other software providers) and deleting the media service providers exception.

However, the proposed Article 88b has been met with considerable resistance and criticism from industry stakeholders[174] while more carefully welcomed in literature[175]. It represents a key aspect against which the success of the Digital Omnibus should be assessed in pursuing its objective of fostering more user-friendly mechanisms while safeguarding innovation in a data-driven economy. If a fundamental rationale of data protection law is to preserve individuals' control over their personal data, and consent is among the most effective means of empowering data subjects, such control also presupposes that individuals possess a sufficient degree of awareness regarding the processing of their data. In line with this assumption a first objection concerns the legal validity of the mechanism. Browser-level consent signals, as currently conceived, are likely to be incompatible with Article 4(11) GDPR, which requires consent to be freely given, specific, informed, and unambiguous[176]. A browser-level mechanism for consent management that comes into play before the data subject has visited the relevant website, received information about the controller's identity or the relevant processing activities (such as the specific purposes) cannot satisfy these requirements. Article 88b essentially replaces the act (i.e., a statement or a clear affirmative action) by which the data subject "signifies agreement to the processing of personal data relating to him or her" with a centralised browser signal. If a clear affirmative action still exists, the mechanism nonetheless undermines the conditions that make consent valid, starting with its informed and specific nature, so that consent becomes a "blanket choice". Accordingly, Article 88b may not resolve the causes of consent fatigue (i.e., data subjects' limited awareness and, consequently, their lack of actual control). Consent choices are often context-dependent and may legitimately vary depending on the controller, the service offered, or the purposes pursued. A browser-level signal inevitably abstracts from these contextual elements and therefore risks transforming consent into a blanket choice that no longer reflects users' actual preferences (i.e., an informed and specific choice) in a specific processing environment. The provision aims at introducing a simplification for users without equally promoting their awareness and facilitating their informed decisions. In fact, on the contrary, this may increase confusion, given that users express their consent when they don't know the downstream, website-level effects of their choice – a choice that,

[173] EDPB-EDPS (2026b), para. 108.
[174] See Ecommerce Europe (2026); IAB (2026); and Alliance Digitale (2026).
[175] See for example Domínguez de Olazábal, I. (2025). The EU's Digital Omnibus Must Be Rejected by Lawmakers. Here is Why. TechPolicy.press, 3 December, who, in the context of a very critical assessment, sees this mechanism as the silver lining of the Digital Omnibus; equally quite supportive, albeit with some recommendations is BEUC (2026). Protecting EU data and privacy rights in the Digital Omnibus, February. On the contrary, see Höppner & Werle (2026), for more specific critical remarks.
[176] See also EDPB (2020). Guidelines 05/2020 on consent under Regulation 2016/679, 4 May.

depending on the underlying business model, may have very different consequences from website to website.

As noted[177], this scenario may also give rise to legal uncertainty on whether browser--level consent would replace or supplement existing website and app consent mechanisms, given the limited awareness of data subjects and the risk that their browser-level consent is not valid. Controllers would still need to collect valid individual consent through their own interfaces and would therefore; in the latter case, providers would face parallel consent flows, raising risks of inconsistency and operational complexity. Curiously enough, the problem would only concern consent, not their refusal: so, while uncertainty may affect browser-level signals and prompt website operators to keep their banners in place, rejections may be validly processed at the browser level and become binding on website operators. This asymmetry may also result in higher incentives for browser operators to set their defaults towards blanket rejection, regardless of the user's actual preferences for any given service.

Additional concerns regard the expected impact of standardisation under Article 88b(4). Whether standardisation organisations can deliver technically reliable and legally adequate standards within the 48-month implementation period remains uncertain, as illustrated by the difficulties the standardisation community has encountered in comparable processes. The proposal has also met with great scepticism from industry: questions are raised as to whether it will effectively address the cookie fatigue – in view of the continued reliance on consent in Article 88a) and the fact that it is not clear whether this primary consent level will effectively be replaced by the browser consent mechanism in Article 88b). Questions are also raised about the feasibility of developing timely standards for AI High Risk systems under the AI Act. Furthermore, risks of fragmentation and duplication of obligations may arise considering the consent management tools already deployed and widely used by industry[178]. In any case, a browser should not be prevented from defaulting higher privacy protections – such as blocking trackers by default.

Finally, Article 88b has raised discussions on its effects on competition. The market for web browsers is highly concentrated. Concentrating consent signalling infrastructure in the hands of a few actors risks creating the anticompetitive effects that the Digital Markets Act seeks to prevent[179]. Taken together, these concerns are of a different order from the technical refinements identified by the EDPB and EDPS, which are cautiously supportive of Article 88b. At a first glance, the proposed introduction of Article 88b might reduce the burden of site-level consent elicitation for users, but it does so at the cost of weakening well-established rationales underpinning consent under the GDPR. Under these conditions, the proposal is unlikely to alleviate consent fatigue effectively. At the same time, its expected impact on website operators does not appear to entail significant benefits in terms of simplification or greater legal certainty and may, in some respects, increase compliance complexity.

4.4.4. Consent management and privacy signals from an economic perspective

From an economic standpoint, both provisions also operate upstream of the AI-related processing analysed in earlier subsections: by reshaping how personal data are stored on, accessed from, and

[177] Höppner & Werle (2026).
[178] Such as the Transparency and Consent Framework – TCF.
[179] Höppner & Werle (2026).

signalled through user terminal equipment, they affect the data flows on which AI training and operation rely. The economic effects on the advertising ecosystem have been analysed elsewhere[180] ; this subsection focuses on the distributional implications for AI-relevant data flows and on the interaction between Articles 88a–88b and the AI-specific provisions analysed in Section 4.2.

Asymmetric effects on AI training data flows. Article 88b's centralised refusal signal applies formally to all non-essential processing, but its incidence on AI training is uneven. AI systems whose training relies on data collected through logged-in user relationships, on contractual data exchanges, or on publicly available web content are largely unaffected: the consent signal regulates terminal-equipment access, not server-side data flows. AI systems whose training depends on cross-site behavioural traces — including some recommender systems, ad-relevance models, and the ancillary models that feed targeting infrastructure — face a structural reduction in admissible data inputs. The asymmetry interacts with the legitimate-interest basis for AI training proposed in Article 88c GDPR (Section 4.2.1): controllers whose training data survive the browser-level signal can rely more readily on legitimate interest, while controllers whose inputs are interrupted face a compounded constraint — narrower data access *and* a more contestable legitimate-interest assessment.

Distributional implications across the AI ecosystem. The combined effect of Article 88b's reduced data availability and Article 88a's broadened derogations for audience measurement and security purposes favours actors whose business models can internalise both functions: large platforms with proprietary first-party data and substantial in-house analytics infrastructure. SME AI developers, and independent firms that depended on intermediated data sources, are proportionately more exposed[181] . The consequence mirrors a pattern documented in the broader empirical literature on the GDPR: regulatory changes whose stated objectives include support for SME competitiveness can, through the mechanism of fixed compliance costs and asymmetric data access, consolidate the position of large incumbents[182]. The interaction with the bias-detection exception introduced in Article 4a of the Omnibus on AI (Section 4.2.3) is also consequential: SME developers that rely on third-party data to construct training sets representative of EU linguistic and cultural diversity face a compounded set of constraints that the Omnibus addresses unevenly across provisions.

AI-driven consent management as part of the solution. Article 88b's machine-readable signal also opens a complementary line of analysis. As foreshadowed in Section 2.1.5, AI tools may improve the quality of consent expression that the site-level architecture has failed to deliver: automated consent management — currently mandated at the browser layer under Article 88b(6), but extensible to other software providers as the EDPB-EDPS Joint Opinion 2/2026 has recommended — can in principle close the gap between stated and revealed privacy preferences documented in Section 2.1, by relieving users of the per-site cognitive load while preserving the granularity of the underlying preferences. Whether this potential is realised depends on the standards adopted under Article 88b(4): only standards that encode purpose-level preferences rich enough to support meaningful differentiation, rather than collapsing the consent decision into a binary accept-reject, will close the gap. The standardisation process therefore sits at the intersection between the simplification objective of the

[180] Decarolis, F., & Firullo, C. (2026). The Digital Omnibus and the Economic Effects of Centralising Consent. *Clifford Chance*.
[181] Demirer et al. (2024); Frey & Presidente (2024).
[182] Jia et al. (2021); Yun (2024).

Digital Omnibus and the broader hypothesis that AI technologies, properly deployed, can reinforce rather than undermine data protection outcomes.

Aggregate assessment. Articles 88a and 88b are economically consequential for the AI-related provisions of the Digital Omnibus despite not being framed as AI-specific measures. Their incidence interacts with the legitimate-interest basis for AI training (Section 4.2.1), with the special-categories regime (Section 4.2.2), and with the data-governance framework discussed in Section 4.5. A simplification of the consent architecture that does not jointly address these interactions risks delivering its efficiency gains unevenly across the European AI ecosystem — favouring actors with proprietary data infrastructures and disadvantaging the SME developers whose growth the Digital Omnibus, the AI Act, and the broader Draghi[183] and Letta[184] agendas have explicitly identified as a priority. Refining the standards under Article 88b(4) and the derogations under Article 88a(3), with explicit attention to their distributional incidence on AI-relevant data flows, is the principal economic question that the negotiation phase will need to address.

[183] Draghi (2024).
[184] Letta (2024).

Part III

5. Conclusion and recommendations

5.1 Overall assessment

The Digital Omnibus enters a debate that this paper has framed in deliberately non-adversarial terms: data protection and AI-driven innovation are not inherently in conflict. The challenge ahead is not to choose between data protection and innovation but to distinguish between scenarios in which those values converge and scenarios in which they genuinely diverge. It is against that standard – not against a presumption that more data protection is always better or that more data access is always worse – that the provisions of the Digital Omnibus must be evaluated for their prospective impact.

The assessment that emerges from Part II of the research is partial and uneven. Three genuine gains from the Omnibus proposal stand out. The codification of legitimate interest as the applicable legal basis for AI model development in Article 88c removes one significant layer of the legal uncertainty that has hindered European AI development and confirms that consent is not the only legal basis for large-scale AI training. Expanding the circumstances in which data can be used to address bias in AI systems is a good idea: the evidence shows that algorithmic bias is not confined to high-risk contexts. And the shift from the ePrivacy Directive to the GDPR in the governance of the storing of, or the gaining of access to personal data, in a terminal equipment is a promising step, although important refinements are still needed.

At the same time, the analysis reveals a recurring pattern of incomplete execution. For example, aspects such as the undefined notion of “operation” of AI systems and the tension between the “unconditional” right to object and Article 21 GDPR, introduce new interpretive uncertainty. Furthermore, uncertainty and cost continues to fall disproportionately on SMEs. Another point of tension lies in the enforcement of data subject rights. The forthcoming EDPB guidelines on the interplay between the GDPR and the AI Act should address this point in a more direct way providing supportive guidance to AI developers.

Finally, an instructive episode concerns an amendment that will not take place: the rejection of the proposed new definition of personal data confirms that tensions between data protection and digital technologies are not productively resolved by narrowing the material scope of the GDPR, but require a more refined articulation of existing mechanisms in light of the peculiarities of AI technologies.

5.2 General recommendations

First, at the level of the Digital Omnibus proposal, priority should be given to clarifying and refining the proposed amendments in order to reduce legal uncertainty without altering the core structure of the GDPR. In this respect, several targeted adjustments appear necessary. The proposed Article 88c should be further specified to reduce ambiguity (notably as regards the meaning of “where appropriate” and the scope of “operation” of AI systems), while ensuring that the legitimate interest basis remains operational for AI training. At the same time, the amendments to Article 9 should explicitly recognise the existence of residual or incidental processing of special categories of personal

data and frame the "avoid–remove–protect" logic as a lifecycle-oriented obligation, rather than a rigid sequence. More generally, the ongoing reforms should acknowledge the technical limits affecting the enforcement of certain data subject rights — particularly rectification, erasure and access — in the context of AI models and ensure that compliance obligations remain grounded in what is technically feasible, without adding further red tape and unnecessary administrative costs.

Second, in the context of the Digital Fitness Check, a more systematic effort should be undertaken to ensure coherence across the EU digital acquis. The interaction between the GDPR and the AI Act currently generates overlaps and potential inconsistencies, particularly with regard to risk management and data governance mechanisms. Clarification is needed on the allocation of responsibilities across multiple actors involved in the AI lifecycle, as well as on the role of technical safeguards – including privacy-enhancing technologies and output filtering – in meeting data protection requirements. In addition, the Fitness Check should reassess the effectiveness of data subject rights as the primary enforcement mechanism in data-intensive technologies.

Third, in a longer-term perspective, targeted adaptation of the GDPR—supported by harmonised guidance rather than wholesale reform, as emphasised in Decarolis and Firullo[185] —may be required to ensure its continued effectiveness in an AI-driven data economy. The current framework, which relies heavily on a combination of principles such as purpose limitation and data minimisation, does not adequately distinguish between scenarios in which data processing serves the interests of both users and developers and those in which it creates genuine risks to fundamental rights. A more future-proof approach would move towards a differentiated, risk-based model, capable of aligning obligations and requirements with the actual risks posed by processing activities. This would entail, inter alia, reconsidering the logic of prohibition vs. derogation for special categories of personal data, reassessing the role of consent and individual control mechanisms, and placing greater emphasis on ex-ante design obligations and accountability mechanisms. Such a shift would not weaken the protection of fundamental rights; rather, it would enhance it by ensuring that regulatory intervention is more targeted and, thus, more effective in addressing the risks that matter most.

5.3 Specific recommendations

The specific recommendations set out below identify the most pressing issues that should be addressed in the ongoing discussions surrounding the Digital Omnibus proposals and, where appropriate, inform the broader reflection taking place in the context of the Digital Fitness Check. While some recommendations concern the refinement of the amendments currently under consideration, others point to structural challenges that cannot be fully resolved through the Omnibus exercise alone and therefore require a more comprehensive reassessment of the EU digital acquis. The recommendations should thus be read as complementary and mutually reinforcing elements of a broader strategy aimed at ensuring that data protection law remains both effective and innovation-friendly in the age of AI.

Preserve legitimate interest as a viable legal basis for AI development. Large-scale AI training often takes place in circumstances where obtaining valid consent is impracticable. The growing recognition by regulators and supervisory authorities that legitimate interest may constitute a suitable legal basis represents an important step towards a more realistic and proportionate application of data

[185] Decarolis and Firullo (2025).

protection law to AI. The challenge is not to choose between innovation and data protection, but to ensure that data protection rules continue to function as safeguards governing the responsible use of data rather than as obstacles to it. This balance is fundamental if Europe is to foster a competitive AI market. As one of the European Union's major achievements, the GDPR should provide a framework for the responsible development and deployment of AI systems that can generate significant benefits for individuals, businesses and society at large, rather than creating unintended barriers to innovation and beneficial use of data. At the same time, reliance on legitimate interest should be accompanied by greater clarification regarding its application in the AI context. In particular, guidance is needed on the distinction between development and operation of AI systems (and the respective roles of the actors involved) and the application of certain principles (transparency, data minimisation, accuracy) in large-scale training activities, including in relation to the exercise of data subject rights. While EU law can establish exceptions to the use of legitimate interest, Member States should not be allowed to impose additional consent requirements, given the resulting risk of fragmentation and the disproportionate burden this would place, e.g., on SMEs and independent developers.

Explicitly recognise the reality of residual and incidental processing of special categories of data. AI models are trained on large and diverse datasets in which special categories of personal data may appear unintentionally or incidentally. A framework that assumes the complete exclusion of such data in all circumstances risks becoming disconnected from technical reality and encouraging merely formal instead of effective compliance. Future-proof rules should acknowledge the existence of residual and incidental processing while maintaining robust safeguards. The objective should not be the elimination of every possible instance of processing of special categories of data, but the implementation of proportionate measures to avoid unnecessary collection, remove such data where feasible and proportionate, and protect any residual information through proper technical guardrails. This life-cycle approach should also inform the operationalisation of data subject rights, particularly rectification and erasure, in the context of AI systems.

Enable bias detection and mitigation through carefully designed exceptions. The use of special categories of data for bias detection and mitigation can serve substantial public interest objectives, such as the prevention of discrimination, while also contributing to the development of fairer AI systems. The upcoming modification of the AI Act is a welcome step towards facilitating the development of more representative and less discriminatory AI systems without creating a general derogation from the protections for sensitive data. The extension of this rule beyond high-risk AI systems and beyond AI providers reflects the awareness that bias detection and mitigation may be necessary across different categories of AI systems and at different stages of the AI lifecycle.

Ensure that the enforcement of data subject rights remains effective and technically realistic. The effectiveness of data protection law in preserving individuals' control over their personal data depends on the exercise of data subject rights. However, the application of rights such as erasure, rectification and access in the context of AI systems raises genuinely new challenges. Future guidance should recognise the technical characteristics of AI models and avoid imposing obligations that are impossible or disproportionate to implement. Instead of technical solutions that, although theoretically feasible, would be impracticable in practice (such as full AI model retraining), equivalent measures delivering comparable protection should be considered. The objective of preserving effective protection for data subjects should not undermine legal certainty or create unnecessary barriers to innovation.

Preserve the right of access as a safeguard while preventing abusive practices. The right of access plays an increasingly important role in enabling individuals to understand and contest the use of AI systems that affect them. In complex and highly automated environments, it contributes not only to transparency but also to accountability. Limitations designed to address manifestly unfounded excessive or abusive requests may be justified. However, such limitations should be applied cautiously in light of the contestation-enabling function of the right of access within the system of data subject rights.

Adopt a more flexible approach to consent to effectively alleviate cookie fatigue. The move towards a more harmonised consent framework for access to and use of terminal equipment is a welcome development. However, a genuinely risk-based approach requires broader exemptions for low-risk processing activities that raise limited privacy concerns, which would equally reduce compliance burdens and consent fatigue. At the same time, the ability of consent to reflect changes in the processing context and users' preferences over time should be preserved, without unduly restricting opportunities to revisit consent choices. Restrictions on renewed consent requests should remain an exception. In particular, fixed limitations such as the proposed six-month ban risk reducing flexibility without improving user autonomy and should therefore be reconsidered.

Preserve users' choice in automated consent mechanisms. The machine-readable privacy signals introduced by Article 88b represent a proposal with ambiguous theoretical effects and potentially disruptive consequences on the online ecosystem. By design, such signals can reduce the number of consent prompts users face, allowing a preference to be set once rather than re-solicited at every site, and thereby easing the repeated interruptions that drive consent fatigue. Reducing the frequency of prompts, however, is not the same as improving the quality of consent: there is no empirical evidence that the mechanism would produce better-informed decisions. On the firms' side, the proposal imposes real adjustment costs, which fall disproportionately on SMEs, while larger controllers are better placed to absorb them. The net effect is therefore genuinely indeterminate: the mechanism might reduce one burden (the frequency of consent prompts) while creating another (controllers' adjustment costs) and may redistribute market power toward the actors that control the signalling layer, with systemic implications for competition. For these reasons, the adoption of a mechanism like the one under the proposed Article 88b should be conditioned on a prior assessment of its structural effects and on credible evidence that automated signals genuinely improve the quality of consent, rather than on the assumption that standardisation will resolve these questions after the fact. At a minimum, the legal effects of such signals should be clarified — in particular, their relationship with the requirements of informed, specific and freely given consent under Article 4(11) GDPR — since the mechanism risks generating additional uncertainty rather than alleviating consent fatigue. Simplification should not come at the expense of users' awareness and control, nor should it introduce a parallel consent layer that duplicates existing mechanisms and compounds legal uncertainty. Any future development of automated signals should accordingly preserve the contextual character of consent.

About CERRE

Providing high quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) is a not-for-profit think tank. It promotes robust and consistent regulation in Europe's network, digital industry, and service sectors. CERRE's members are regulatory authorities and companies operating in these sectors, as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach covering a variety of markets, such as energy, mobility, sustainability, tech, media, telecom, etc.;
- the widely acknowledged academic credentials and policy experience of its research team and associated staff members;
- its scientific independence and impartiality, and;
- the direct relevance and timeliness of its contributions to the policy and regulatory development process impacting network industry players and the markets for their goods and services.

CERRE's activities include contributions to the development of norms, standards, and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological, and social environment. CERRE's work also aims to clarify the respective roles of market operators, governments, and regulatory authorities, as well as contribute to the enhancement of those organisations' expertise in addressing regulatory issues of relevance to their activities.

About the Authors

Marco Bassini is an Assistant Professor of Fundamental Rights and Artificial Intelligence at the Tilburg Institute for Law, Technology, and Society – Tilburg University, a Research Fellow at CERRE and a Fellow at the Baffi Centre on Economics, Finance and Regulation – Bocconi University. Previously, he served as an Adjunct Professor of Constitutional Law and Internet Law at Bocconi University, where he also was the coordinator for the LLM programme in Law of Internet Technology from 2020 to 2022. In 2016 he obtained his PhD in Constitutional Law and European Law from the University of Verona. For over a decade, he has combined his academic career with legal practice, working as a qualified attorney and serving as a data protection officer with leading organisations. He also served as external adviser to the Italian Communications Authority and the Italian Ministry for Technological Innovation. His research interests include the protection of human rights in the digital age, AI regulation, platform regulation.

Cristiana Firullo is a PhD Candidate at the Department of Information Science at Cornell University and Visiting Scholar at the Department of Economics at Stanford University. Her research lies at the intersection of economics, behavioural science, and digital platforms, focusing on online user behaviour and digital market design, in particular advertising, algorithmic recommendation systems, and data-driven intermediaries. She combines field experiments, large-scale behavioural data, and theoretical modelling to study how platform design choices affect consumer welfare, market outcomes, and privacy. She is a member of the Internet Behavior Experiment (IBE), a large-scale randomized field experiment on online advertising, tracking, and targeting across browsers, email, and mobile devices. She holds a BSc and MSc in Economics and Social Sciences from Bocconi University, and previously worked at Compass Lexecon (London) and Oxera Consulting LLP (Oxford).

cerre

Avenue Louise 475 (box 10)
1050 Brussels, Belgium
+32 2 230 83 60
info@cerre.eu
www.cerre.eu

Centre on Regulation in Europe (CERRE)
CERRE Think Tank
CERRE Think Tank