OECD
BETTER POLICIES FOR BETTER LIVES

Policy Brief

Enhancing user control and privacy protections in the context of AI

October 2026

Key messages

- Building AI machine-learning models often requires large volumes of high-quality data, which may (purposefully or inadvertently) include personal data. This creates a dual need: to continue facilitating access to high-quality data across the AI lifecycle, while ensuring privacy and users' control over their personal data. Both objectives are complementary since privacy-compliant data collection contributes to data quality by fostering trust and more willingness to share data. These objectives also need to be pursued alongside appropriate safeguards for other legitimate rights and interests involved in the collection, access and use of data, including intellectual property rights (IPRs) and other proprietary rights.
- A significant share of AI training data originates from user interactions with generative AI systems, particularly in business-to-consumer (B2C) settings such as AI chatbots, as well as from the scraping of data that is publicly available on the internet. In addition to static training datasets, AI systems are increasingly integrating dynamic approaches such as Retrieval-Augmented Generation (RAG) to access and incorporate external information in real time.
- Recent regulatory guidance from Privacy Enforcement Authorities (PEAs) emphasises that AI chatbots should inform users about how personal data from prompts are processed, and whether they will be used for training or refining models. Many developers are improving transparency and offering global opt-out mechanisms, but there is scope to design more granular user controls, supported by "just-in-time transparency" mechanisms that deliver relevant information at the moment of interaction rather than relying solely on privacy policies.
- Personal data can become embedded in AI model weights or "memorised" during training, creating challenges for access, correction, and erasure. Fully removing data from such models would require significant technical effort, calling for pragmatic collaboration between industry and regulators to protect user interests while taking technical realities into account.
- As AI systems evolve towards more autonomous agentic forms, the focus will shift from direct user interactions to agentic/multi-agent systems that can take complex actions autonomously. This will challenge the application of traditional data protection principles, but AI agents could also support users in managing privacy across complex data interactions.
- The governance of web-scraped data for AI training is evolving through a combination of regulatory guidance, industry standards and technical safeguards.

A growing need to access high-quality and privacy-compliant data to support trustworthy AI

Data have played a central role in propelling AI machine-learning models forward over the last decades. In addition to scaling up model size and computational power, the use of larger AI training datasets has been a significant driver of AI performance improvements. Reflecting this trend, recent OECD work has underscored that the volume of data used to train AI models has been increasing since 2010, growing by 2.6 times each year (Hobbs et al., 2026[1]). While specific data needs vary depending on the type and purpose of the model, greater access to accurate, relevant, and high-quality data generally enables AI models to perform better, as they have the ability to learn from data points and patterns inferred in an iterative process (OECD, 2022[2]).

The growing demand for large volumes of training data can create tensions with privacy and data protection requirements (OECD, 2025[3]). As organisations seek to access and share more data for AI development, concerns may arise regarding lawful data collection, the protection of personal information, and individuals' control over their data. Nevertheless, these tensions do not necessarily amount to a trade-off. Exploring synergies between the AI and privacy communities, as advocated in the report *"AI, Data Governance and Privacy: Synergies and Areas of International Co-operation"* (OECD, 2024[4]), can support a more nuanced policy approach that enables access to and sharing of data for AI training – both in terms of volume and quality – while safeguarding other legitimate rights and interests, including privacy, user control over personal data, IPRs, and legitimate commercial and non-commercial interests. This approach is consistent with the OECD Recommendation on Enhancing Access to and Sharing of Data (EASD) (OECD, 2021[5]), which sets out principles for maximising the benefits of data access and sharing while protecting the rights of individuals and organisations, and which are particularly useful in the context of AI (OECD, 2025[6]).

At the same time, recent research suggests that the pace of AI training scale-up may be constrained by data limitations, as the annual growth in human-generated publicly available text data is decelerating compared to training data demands (Villalobos et al., 2024[7]). In this scenario, rather than algorithms necessarily stagnating, the availability of data itself could become an increasingly important bottleneck for developing more capable AI systems.

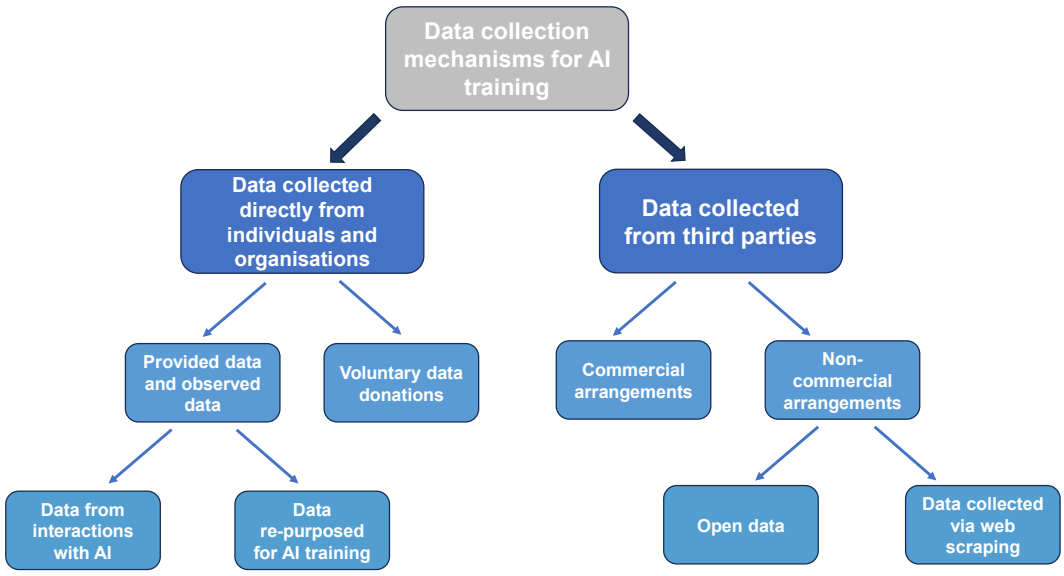
But scaling models alone does not guarantee improved performance: large language models (LLMs) may sometimes exhibit inverse scaling, where performance declines as models grow larger due to flaws or low quality in their training data (McKenzie et al., 2024[8]). From a "data perspective", therefore, continuing to improve AI performance will require not only more data, but also more diverse and higher-quality datasets (OECD, 2025[9]) that can help address algorithmic biases, reduce errors, and limit unintended outcomes in AI systems, thereby contributing to their overall trustworthiness. Ensuring such diversity and quality may also require addressing structural disparities in data availability and representation across countries and contexts.

Importantly, the characteristics of high-quality data (data which are accurate, complete, consistent, reliable, valid, timely) closely align with common data protection principles (data quality, data minimisation, purpose principle, etc.). Moreover, national frameworks for AI governance increasingly recognise that lawful, well-governed, reliable and representative data are important elements of data quality and trustworthy AI (UK Government, 2026[10]), (MIC and METI, 2026[11]), (Australian Government, 2024[12]), (IMDA and PDPC, 2022[13]). As well, innovation in privacy technologies such as synthetic data and federated learning, the development of technical safeguards – such as operational guardrails and filters – at model input and output stages, the growing availability of privacy dashboards, etc., contribute to such alignment, despite undeniable technical and regulatory challenges (OECD, 2025[14]).

Against this backdrop, understanding how AI developers collect and access data in practice becomes essential for understanding how the demand for larger and higher-quality datasets can be reconciled with transparency, user control, and the protection of privacy rights. The paper "*Mapping Relevant Data*

Collection Mechanisms for AI Training" (OECD, 2025[9]) provides a useful framework in this regard. By mapping the key mechanisms currently used by AI developers to source AI training data, it helps clarify where potential privacy risks may arise, where safeguards can be embedded, and how certain approaches may better support user agency, privacy controls, and data protection rights.

Figure 1. Key data collection mechanisms for AI training



Note: This figure presents an overview of the key data collection mechanisms for AI training highlighted in the document. It does not necessarily capture all of the different ways of collecting data for AI training – the AI ecosystem is dynamic, and new mechanisms and sources will likely evolve as new business players and models enter the AI field
Source: Adapted from (OECD, 2025[9]), "Mapping relevant data collection mechanisms for AI training", OECD Artificial Intelligence Papers, No. 48, OECD Publishing, Paris, https://doi.org/10.1787/3264cd4c-en.

The taxonomy focuses on relevant data collection mechanisms used for both the training and fine-tuning phases of the AI model's development. By systematically categorising how AI developers source their data, it provides a concrete basis for identifying where tensions between data access and privacy protection are most likely to emerge, and where targeted policy responses and technical safeguards can be most effectively applied. It first differentiates based on the sources[1] from which AI developers obtain data: i) directly from individuals and organisations; and ii) from third-party sources and providers. Within these two primary data sources, it identifies different relevant data collection mechanisms outlined in the figure above.

In addition to the key data collection mechanisms shown in Figure 1, novel methodologies bridge the training and deployment stages by enabling dynamic data access. For instance, Retrieval-Augmented Generation (RAG) enables LLMs to access and incorporate data from external sources when responding to user queries in human-AI interactions. As a result, AI systems can draw not only on data included in the original pre-training corpus, but also up-to-date information from sources such as websites, raising privacy considerations that training-time safeguards alone cannot fully address.

In light of these developments, growing policy and industry attention is being brought to strengthening user control and privacy protection across the AI lifecycle. As discussed at the GPAI workshop "User control

and privacy rights in relation to AI systems" (hereinafter "GPAI workshop") (OECD, 2025[3]), this concern stems from the recognition that legal compliance, trust, and data accuracy are key to the quality of large-scale data use because they include meaningful user control and oversight. In this context, the following sections focus on two prominent data collection mechanisms identified in the taxonomy: data collected directly from users during their interactions with generative AI systems such as chatbot interfaces, and data obtained through web scraping, without direct engagement with the individuals whose data are collected.

How to ensure user control over personal data when users interact with generative AI systems?

As previously documented by the OECD, a growing volume of training data originates from data submitted by users or passively collected during their interactions with generative AI systems, especially business-to-consumer (B2C) tools like chatbots (OECD, 2025[9]). Indeed, there has been a rapid development and use of AI chatbots worldwide,[2] including for virtual friendships and therapeutic purposes. When interacting with AI chatbots, users may end up sharing more personal data than they intend. The conversational nature of AI chatbots encourages users to be more candid, sharing personal data and thoughts more freely and naturally than with traditional digital channels (Croes et al., 2024[14]) and (Zhang et al., 2024[15]). Unlike typical web browsing or social media activity, exchanges with AI tools frequently capture users' raw thoughts, spontaneous reflections, and sometimes emotionally charged reflections. These risks are particularly pronounced for generative AI chatbots, which, unlike rule-based systems, can memorise and in some cases inadvertently reproduce personal data encountered during training – a challenge discussed further below. Additionally, users may be unaware that their conversations are being logged or used for AI training purposes. This can create a disconnect between user expectations and actual data practices involved in AI training (OECD, 2024[4]).

Strengthening privacy safeguards and user control over data collection practices are therefore essential to build public trust in the protection of users that interact with AI systems and whose data contributes to AI training. Achieving this in practice requires a mix of clear disclosure when personal data are collected or used for AI training, mechanisms that effectively allow users to exercise control over their personal data, interfaces designed to support meaningful user choice, and robust privacy safeguards across the AI lifecycle. While existing privacy and data protection frameworks in most jurisdictions already provide a foundation for user protections over personal data, applying these requirements effectively in AI conversational settings is complex and has prompted the technical and regulatory developments discussed below.

A pragmatic regulator-industry collaboration to strengthen user control and rights

Transparency and user opt-out mechanisms

A growing number of Privacy Enforcement Authorities (PEAs) have issued specific guidance on how to address privacy risks when users interact with AI systems, and more specifically with AI chatbots. Guidance commonly recommends the implementation of transparency measures and privacy safeguards such as data minimisation or anonymisation.

For example, the Office of the Privacy Commissioner of Canada's (OPC) *Principles for responsible, trustworthy and privacy-protective generative AI technologies* (OPC, 2023[16]) highlights that developers and providers of generative AI systems should inform any users interacting with a generative AI system about where personal data collected from prompts are used for further training or refining of an AI model. In its *AI and algorithmic risk report* (Autoriteit Persoonsgegevens, 2025[17]), the Dutch Data Protection Authority also emphasises that AI chatbots must be transparent regarding the processing of personal data,

the types of data involved, and the potential purposes for which the data are used. Similarly, the Office of the Australian Information Commissioner's (OAIC) *Guidance on privacy and developing and training generative AI models* (OAIC, 2024[18]) clarifies that AI developers must provide clear notice to users if they intend to use prompts and chatbot responses for training future versions of the foundational chatbot model. The OAIC recommends placing a brief statement above the input field informing users that their interactions might be leveraged for model re-training, with a more detailed explanation in the privacy policy. Singapore's Personal Data Protection Commission (PDPC), in its *Proposed Advisory Guidelines on Use of Personal Data in Generative AI* (PDPC, 2026[19]), recommends that organisations explicitly inform users when their personal data will be used to develop or improve AI or generative AI models, rather than relying on general notifications.

Industry actors are responding with complementary technical and operational measures to increase transparency. Many providers of widely used AI applications, such as AI chatbots, provide information about whether and how user prompts and conversations may be used to improve or train their models and have introduced mechanisms that allow users to limit or opt-out of such uses. These control mechanisms are still evolving, and approaches vary across companies. For example, several developers offer settings that allow users to opt-out of having their conversations used for model training or improvement, typically through in-app controls or account-level activity settings, as documented in transparency reports voluntarily submitted under the Hiroshima AI Process (HAIP) Reporting Framework (OECD, 2025[20]).

These controls generally apply across all user activity, meaning that users cannot selectively exclude specific conversations, files, or categories of information from being used for AI training. Because AI chatbots operate through iterative exchanges, the nature and sensitivity of users' input data may vary from one interaction to another. This creates an opportunity to consider more differentiated control models that could better reflect the contextual nature of user interactions. Potential approaches could include more granular management of data use at the level of individual conversations or categories of data (OECD, 2024[4]). In order to increase user control over personal data, such measures could be supported by "just-in-time transparency" mechanisms, giving users relevant information at the moment of interaction with the chatbot – such as a short notification above the input field before sensitive exchanges – rather than relegating disclosures to help pages or privacy policies (OECD, 2025[3]). Such approaches are particularly relevant where users may share sensitive personal data, including children and other vulnerable users, and can help reduce the gap between user expectations and actual data practices. The effectiveness of such mechanisms also depends on whether users can understand the privacy-related information they receive and can effectively act on it. This is particularly relevant in contexts with varying levels of digital literacy, where interface design and default settings play an important role in enabling meaningful user control.

Privacy-enhancing technologies (PETs) and other technical safeguards

Regulators also highlight the need to proactively address privacy risks associated with AI interactions, such as heightened risk of data breaches where users share sensitive information without realising that it may be stored or incorporated into training datasets (OAIC, 2024[18]). The OPC recommends using anonymised or de-identified information in prompts wherever possible and avoiding the retention or secondary use of prompts containing personal data unless the user gives permission (OPC, 2023[16]). Similarly, Singapore's PDPC advocates the use of anonymised or pseudonymised data where possible, and provides guidance on considerations organisations should weigh when using such data in place of personal data (PDPC, 2024[21]). Korea's Personal Information Protection Commission (PIPC), in its *Guidelines on the Processing of Personal Information for the Development and Use of Generative AI*, advises implementing input and output filtering mechanisms and applying enhanced safeguards when AI systems interface with external databases, for example, through RAG techniques (PIPC, 2026[22]).

PETs applied at the training stage are complementary safeguards that industry is exploring. Differential privacy can prevent individual conversations from being identified, while federated learning enables model improvement without centralising raw conversation data (OECD, 2025[23]). Contextual integrity frameworks (which are meant to ensure that information flows respect the context in which data was originally shared) and zero-knowledge proofs (which can allow verification of data handling without exposing the underlying data itself) offer further promising avenues for enhancing user control and safeguarding sensitive data (OECD, 2025[3]). PETs are best understood as part of a broader governance architecture, rather than a substitute for legal frameworks (OECD, 2023[24]). Their value also depends on appropriate implementation, evaluation and documentation, as well as adequate institutional capacity and practical deployment support, which may vary across jurisdictions.

Beyond mitigating privacy risks, PETs are also explored by industry to enable the trusted use and sharing of data for AI development and collaboration. Recent guidance from Singapore's PDPC highlights how PETs are evolving from compliance tools into strategic enablers of responsible data use, supported by practical implementation resources such as PET sandboxes, discovery tools and implementation checklists (PDPC, 2025[25]).

Emerging industry practice also involves using small language models fine-tuned as "compliance detectors" to identify privacy-relevant content in user interactions (OECD, 2025[3]). As this landscape of technical interventions continues to mature and diversify, greater consistency in the use of terms such as guardrails, filters, and safeguards across industry practice could support comparability, effective oversight, and policy coordination. As discussed at the GPAI workshop, the development of taxonomies would be a meaningful step in this direction, as shared vocabulary would help regulators and users understand the protections in place, and enable organisations to benchmark their privacy measures against industry standards and regulatory expectations.

Mechanisms enabling user rights and redress

PEAs in various jurisdictions have also highlighted the importance of users having effective avenues to address the use of their personal data across both the input and output phases of AI systems. This includes mechanisms that allow them to request access to, correction or deletion of personal data, as well as processes to challenge or seek the removal of content generated using their personal data (OECD, 2025[3]). Regulatory practice across jurisdictions is beginning to give these principles more concrete form. The UK Information Commissioner's Office (ICO) has clarified that data subject rights apply throughout the AI lifecycle and that practical implementation challenges cannot justify non-compliance (ICO, 2026[26]). Canada's OPC has similarly highlighted the importance of maintaining records and traceability for AI-assisted decisions in order to enable individuals to exercise access rights and challenge decisions involving their personal data (OPC, 2023[16]). The exercise of rights in a domestic context effectively depends on individuals being able to identify the relevant data controller and direct their requests accordingly, underscoring the importance of transparency around roles and responsibilities in AI systems where these may be distributed across multiple actors.

These principles are tested by risks such as membership and property inference attacks, and hallucinations that may generate plausible but false personal information about individuals, raising questions about the scope of rectification and redress mechanisms available to users. These complexities include emerging questions around whether correcting personal data used in AI systems may, in some cases, extend to the model itself. Experts at the GPAI workshop debated whether such correction could, in some circumstances, involve "unlearning" information from AI models, and whether the technical challenges of doing so could prompt discussions about new or adapted legal and policy approaches (OECD, 2025[3]). Some emerging guidance suggests that, where personal data are embedded or memorised within an AI model, rectification obligations may apply not only to training data or outputs but also to the model itself (CNIL, 2026[27]) (ICO, 2026[26]). These evolving approaches to data subject rights

are mirrored by broader international efforts to strengthen user control over AI-generated content. A recent Joint statement on AI-generated imagery calls for organisations to provide users with effective and accessible mechanisms to request the rapid removal of harmful or non-consensual content involving their personal data (Joint statement by data protection and privacy authorities, 2026[28]).

User access to capabilities in this area are improving, with privacy dashboards increasingly providing users with insights into inferred data, AI profiling, and personal data included in model outputs. Expert discussion at the GPAI workshop also surfaced the proposal to give users access to their interaction logs in a format that could be transformed into personal evaluation datasets, enabling individuals to audit how their data have shaped model behaviour and protect their privacy more meaningfully (OECD, 2025[3]). Nonetheless, fulfilling these rights at scale remains technically and operationally challenging. Regulators are increasingly acknowledging these technical limits, marking a shift towards more pragmatic collaboration with industry. Tools such as policy engines, which automate the application of privacy rules at the system level, illustrate how compliance is increasingly becoming a technical function embedded in AI infrastructure – a development that both enables and requires closer coordination between regulatory expectations and engineering practice (OECD, 2025[3]).

Collaboration between regulators and industry is particularly evident in efforts to address the technical challenge of data memorisation. A persistent tension arises when personal data have been integrated into the AI model in such a way that it becomes part of its learned representation (OECD, 2024[4]). While removing data from the back-end training database is often sufficient to prevent adverse effects, when an AI system has "memorised" personal data during training, there is a risk of "regurgitation" in the sense that such data could still be reproduced or extracted from the model itself. This raises questions about the application of relevant data protection principles, including data minimisation and the right to erasure in various jurisdictions, and illustrates the practical challenges of giving effect to these rights in these jurisdictions once personal data have become embedded in a model.

However, fully removing personal data from an AI model requires significant technical effort or, in some cases, may not be possible without destroying the model itself (OECD, 2025[3]). Expert discussions at the GPAI workshop underscored a continuing stalemate: while full retraining would be the most reliable way to ensure personal data are removed, it is often unfeasible for large general-purpose models. In response, regulatory discourse is gradually evolving towards more operationally proportionate approaches, and technological innovation has given rise to methods such as machine unlearning, which aim to remove specific data points from the model without the need for complete retraining (OECD, 2025[3]). While research on machine unlearning techniques is still evolving and current approaches remain technically challenging to implement and evaluate at scale, they offer promising pathways to support data erasure while easing the operational and economic burden on model developers.

This memorisation challenge is particularly acute for AI chatbots built on open-source models, which are routinely fine-tuned and redistributed across platforms. In France, the *Commission nationale de l'informatique et des libertés* (CNIL) has developed a model genealogy tool to trace these lineages, helping compliance teams to identify model versions that may contain memorised personal data from prior chatbot interactions and respond to access or deletion requests (CNIL, 2026[29]). Beyond traceability, the practical distribution of responsibility across the chain of downstream fine-tuners remains a key consideration for the effective enforcement of data erasure, as fulfilling a data deletion request at the source does not automatically reach every redistributed version of the model.

Future outlook: AI agents

Looking ahead, as AI systems evolve beyond conversational interfaces towards more autonomous forms, regulators and industry will need to find avenues of collaboration around the application of user control principles to AI agents[3] capable of taking actions on behalf of users with a degree of autonomy (OECD, 2026[30]). These systems may challenge the application of existing data protection concepts (ICO, 2026[31]),

such as consent and purpose limitation, designed for direct user interactions. Consent is particularly at risk of degrading across agent-to-agent handoffs, as adherence to the scope and purpose of data use originally authorised by the user becomes harder to enforce at each successive step. Agentic systems also introduce new risks such as the autonomous inference or exposure of confidential information across services, and the reuse of personal data entered in prompts in ways that users may not anticipate or control. Enforcing data deletion will be particularly challenging in agentic/multi-agent systems, where personal data flows across models, tools, and orchestration layers. At the same time, these advancements may create opportunities, for example if AI agents can act as privacy-supporting tools that help users enforce their own preferences, apply automated safeguards, and manage complex data interactions across services (OECD, 2025[3]) (ICO, 2026[31]). Future work could therefore explore both data governance challenges and AI agents' potential to strengthen user control, potentially through a use-case analysis.

Web scraping: a multi-level approach to manage data privacy issues

The growing popularity of data scraping as a data collection tool has led to widespread use for AI model training, particularly LLMs. The volume of traffic originating from scraping bots for AI training remains substantial, followed by traffic from AI search bots that retrieve content using RAG to support AI chatbot interactions (Cloudflare, 2026[32]). Because scraped data are frequently used at the pre-training stage and subsequently combined with internal or user-derived data during fine-tuning, privacy risks introduced through scraping – such as the inclusion of personal data later memorised by the model – can resurface in downstream applications like AI chatbots, reinforcing the risks discussed earlier in this brief.

OECD analysis of IPRs considerations is presented in the report (OECD, 2025[33]) "*Intellectual property issues in artificial intelligence trained on scraped data*". Data scraping practices also raise concerns from the privacy and data protection perspective (Joint statement by data protection and privacy authorities, 2024[34]). A key policy question is how personal data can be lawfully collected through web scraping while respecting applicable legal rights. Policy and industry efforts have increasingly sought to achieve this balance, focusing on governance frameworks and technical privacy safeguards that enable users to exercise meaningful control over their personal data.

How are regulators and industry addressing user control and rights in data scraping practices?

PEAs in various jurisdictions have issued guidance addressing how privacy and data protection law applies to data collected by web scraping. For example, the European Data Protection Board's ChatGPT Taskforce highlighted that developers relying on the legal basis of legitimate interest – provided they can demonstrate compliance with the GDPR three-part balancing test – should implement safeguards such as defining precise and proportionate data collection criteria, excluding special categories of data, avoiding certain high-risk sources (such as public social media profiles), and applying filtering or anonymisation measures before the training stage (EDPB, 2024[35]). The CNIL also recommends excluding sensitive sources from collection, limiting scraping to clearly publicly accessible data, respecting technical or contractual restrictions on data reuse, implementing mechanisms that allow individuals to object to the collection of their data, and avoiding the use of persistent identifiers unless strictly necessary (CNIL, 2026[36]). Likewise, the UK ICO emphasises transparency as a key safeguard in the context of data scraping, advising organisations to clearly inform individuals about whether their personal data are processed, the sources of the data, and the purposes for which they are used (ICO, 2025[37]). Beyond the European context, the Concluding joint statement on data scraping and the protection of privacy (Joint statement by data protection and privacy authorities, 2024[34]) reflects the signatories' view that "social media companies and the operators of websites that host publicly accessible personal data have an obligation to protect publicly

accessible personal data on their platforms from data scraping that violates data protection and privacy laws".

At industry level, organisations have developed a range of external controls designed to prevent unlawful scraping at source. Many websites and platforms explicitly prohibit automated data collection through their terms of service, creating potential contractual violations when scraping activities occur without permission, even if the data are publicly viewable (Longpre et al., 2024[38]). In addition to terms of service, websites have increasingly adopted technical safeguards (such as "robots.txt" files)[4] to restrict data scraping, especially since mid-2023. However, the limited effectiveness of protocols like robots.txt, which rely on voluntary bot compliance rather than technical enforcement, has prompted the development of stronger access control tools. One example is *Cloudflare's AI Scrapers and Crawlers* feature, which allows website owners to block AI bots at the network level with a single setting, without needing to manually configure rules for each bot individually (Cloudflare, 2024[39]). It is also worth noting that oversight mechanisms developed to enhance transparency, traceability and user control over data collection practices draw heavily on instruments primarily designed to protect IPRs. Content identification tools, such as the International Standard Content Code (ISCC) (ISO, 2024[40]), rights-holder opt-out frameworks like Creative Commons CC Signals (Hardinges, Pearson and Ross, 2025[41]), emerging standards initiatives such as the IETF AI Preferences Working Group (IETF, 2025[42]), and machine-readable privacy standards such as the IEEE P7012 ("MyTerms") (IEEE, 2025[43]) are examples of machine-readable governance frameworks designed to signal usage preferences, enable traceability, and support compliance around AI training datasets. The effectiveness of these mechanisms is also contingent on how well they are adopted, enforced and made accessible, including for smaller actors such as independent creators, scientific communities and cultural institutions.

Where data have been collected, internal governance tools provide a second layer of protection. Major AI developers report applying multi-stage filtering and risk-mitigation processes to improve the quality and safety of training datasets, including those derived from publicly available sources (OECD, 2025[20]). As highlighted in the HAIP Reporting Framework submissions, these privacy filters typically rely on machine-learning classifiers, rule-based systems, and metadata tagging to identify and exclude personal data, sensitive content, and material that may pose safety or compliance risks before incorporation into model weights (OECD, 2025[20]). Experts at the GPAI workshop also noted the growing use of data lineage tracking to map how scraped data flows from public sources into training sets. This capability is particularly critical when dealing with heterogeneous web-scraped data from multiple sources, enabling compliance teams to respond to user requests for access, correction, or deletion even after data integration (OECD, 2025[3]). Guardrails and evaluation processes provide a further layer of quality assurance to manage privacy risks associated with scraped data. Developers increasingly conduct red-teaming exercises (including offline testing against static datasets) designed to test vulnerabilities like personal data regurgitation from web-scraped training sets. These exercises help identify weaknesses in both input-stage scraping filters and output safeguards, enabling iterative improvements to protect against unintended personal data disclosure (OECD, 2025[3]).

Transparency and documentation mechanisms complement these technical controls by strengthening accountability and supporting the exercise of user rights. Model cards and data cards can help disclose how datasets were scraped, filtered, and processed from public sources (OECD, 2025[3]). By documenting scraping sources, exclusion criteria, and filtering practices, these mechanisms support privacy audits and enable users to better understand whether and how their publicly available data may have been incorporated into AI training datasets.

These developments reflect how regulators and industry stakeholders are increasingly exploring more collaborative approaches that combine governance frameworks, technical guardrails, and evaluation methods to strike a balance between lawful data scraping and the protection of privacy and data protection interests – particularly given that scraping often occurs without the knowledge of the users whose data are collected. International co-operation has an important role to play in this regard, supporting countries –

especially those with developing regulatory frameworks – exchange practices on guidance, enforcement, documentation standards, and mechanisms that enable user rights and redress.

What can policymakers do to enhance user control over personal data in the context of AI?

- Encourage appropriate and technically grounded regulatory approaches that acknowledge the computational and architectural limitations of AI systems, such as the challenges of identifying personal data in model weights or the complexities of large-scale retraining, while promoting feasible and scalable solutions for data removal and the mitigation of privacy risks.
- Support capacity-building and knowledge-sharing to ensure that evolving technical realities are reflected in policy discussions across jurisdictions with different levels of institutional maturity.
- Promote research, experimentation, and innovation in emerging technologies such as machine unlearning and PETs that enable users to protect their data without undermining the value of AI systems; this includes exploring the potential for AI agents to serve as privacy defenders that dynamically enforce protections and support users in managing complex data interactions.
- Empower the capabilities of everyday users in AI environments, by promoting the design of intuitive, no-code or low-code interfaces that allow them to configure and manage how systems use their personal data.
- Support ongoing practical, risk-based regulatory guidance on how privacy and data protection frameworks apply to personal data collected via web scraping or new data collection techniques, such as retrieval-augmented generation.
- Foster a common understanding and more consistent terminology for privacy safeguards in AI systems – such as guardrails, filters, and compliance mechanisms – to support transparency, comparability, and effective oversight.
- Strengthen dialogue and co-operation between regulators, AI stakeholders, researchers and civil society, ensuring that evolving technical realities are reflected in policy discussions and that mechanisms for user control over personal data remain effective as AI architectures advance.

Further information

OECD (n.d.), Recommendation concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, OECD/LEGAL/0188, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188.

OECD (n.d.) Recommendation on Enhancing Access to and Sharing of Data, OECD/LEGAL/0463, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463.

OECD (n.d.) Recommendation on Artificial Intelligence, OECD/LEGAL/0449, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

OECD (2024), “AI, data governance and privacy: Synergies and areas of international co-operation”, OECD Artificial Intelligence Papers, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/2476b1a4-en.

OECD (2025), “Mapping relevant data collection mechanisms for AI training”, OECD Artificial Intelligence Papers, No. 48, OECD Publishing, Paris, https://doi.org/10.1787/3264cd4c-en.

OECD (2025), “Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence”, OECD Policy Briefs, No. 13, OECD Publishing, Paris, https://doi.org/10.1787/23a70dca-en.

OECD (2025), “Sharing trustworthy AI models with privacy-enhancing technologies”, OECD Artificial Intelligence Papers, No. 38, OECD Publishing, Paris, https://doi.org/10.1787/a266160b-en.

References

Australian Government (2024), *National framework for the assurance of artificial intelligence in government*. [12]

Autoriteit Persoonsgegevens (2025), *AI & Algorithmic Risks Report Netherlands*, https://www.autoriteitpersoonsgegevens.nl/en/current/ap-ai-chatbot-apps-for-friendship-and-mental-health-lack-nuance-and-can-be-harmful (accessed on 19 February 2026). [18]

Cloudflare (2026), *Cloudflare Data Explorer*, https://radar.cloudflare.com/explorer?dataSet=ai.bots&dt=52w&groupBy=user_agent&filters=crawlPurpose%253DTraining (accessed on 15 March 2026). [32]

Cloudflare (2024), *Declare your AIndependence: block AI bots, scrapers and crawlers with a single click*, https://blog.cloudflare.com/declaring-your-aindependence-block-ai-bots-scrapers-and-crawlers-with-a-single-click/ (accessed on 17 February 2026). [39]

CNIL (2026), *AI system development: CNIL’s recommendations to comply with the GDPR*, https://www.cnil.fr/en/ai-system-development-cnils-recommendations-to-comply-gdpr (accessed on 12 March 2026). [27]

CNIL (2026), *La CNIL publie un outil pour la traçabilité des modèles d’IA publiés en source ouverte*, https://www.cnil.fr/fr/la-cnil-publie-un-outil-pour-la-tracabilite-des-modeles-dia-publies-en-source-ouverte (accessed on 6 March 2026). [29]

CNIL (2026), *The legal basis of legitimate interest: focus sheet on the measures to implement in the case of data collection by web scraping*, https://www.cnil.fr/en/legal-basis-legitimate-interest-focus-sheet-measures-implement-case-data-collection-web-scraping (accessed on 18 February 2026). [36]

Croes, V. et al. (2024), “Digital Confessions: The Willingness to Disclose Intimate Information to a Chatbot and its Impact on Emotional Well-Being”, *Interacting with Computers*, Vol. 36/5, pp. 279–292, https://doi.org/10.1093/iwc/iwae016. [15]

EDPB (2024), *Report of the work undertaken by the ChatGPT Taskforce*. [35]

Grand View Horizon (2026), *Global Chatbot Market Size & Outlook, 2025-2030*, https://www.grandviewresearch.com/horizon/outlook/chatbot-market-size/global (accessed on 19 February 2026). [44]

Hardinges, Pearson and Ross (2025), *From Human Content to Machine Data: Introducing CC Signals*, Creative Commons, https://creativecommons.org/wp-content/uploads/2025/06/Human-Content-to-Machine-Data_Final.pdf. [41]

Hobbs, H. et al. (2026), "Exploring possible AI trajectories through 2030", *OECD Artificial Intelligence Papers*, No. 55, OECD Publishing, Paris, https://doi.org/10.1787/cb41117a-en. [1]

ICO (2026), *Guidance on AI and Data Protection – How do we ensure individual rights in our AI systems?*, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/ (accessed on 12 March 2026). [26]

ICO (2026), *ICO tech futures: Agentic AI*, https://ico.org.uk/about-the-ico/research-reports-impact-and-evaluation/research-and-reports/technology-and-innovation/tech-horizons-and-ico-tech-futures/ico-tech-futures-agentic-ai/. [31]

ICO (2025), *The lawful basis for web scraping to train generative AI models*, https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/response-to-the-consultation-series-on-generative-ai/the-lawful-basis-for-web-scraping-to-train-generative-ai-models/ (accessed on 18 February 2026). [37]

IEEE (2025), *IEEE Standard for Machine Readable Personal Privacy Terms*, https://standards.ieee.org/ieee/7012/7192/. [43]

IETF (2025), *AI Preferences Working Group*, https://datatracker.ietf.org/doc/charter-ietf-aipref/. [42]

IMDA and PDPC (2022), *AI Verify – An AI Governance Testing Framework and Toolkit*. [13]

ISO (2024), *Information and documentation - International Standard Content Code (ISCC) (ISO 24138:2024)*, https://www.iso.org/standard/77899.html (accessed on 18 February 2026). [40]

Joint statement by data protection and privacy authorities (2026), *Joint statement on AI-generated imagery and the protection of privacy*, https://www.priv.gc.ca/en/opc-news/speeches-and-statements/2026/s-d_20260223_ai/ (accessed on 2 March 2026). [28]

Joint statement by data protection and privacy authorities (2024), *Concluding joint statement on data scraping and the protection of privacy*, https://www.priv.gc.ca/en/opc-news/speeches-and-statements/2024/js-dc_20241028/ (accessed on 17 February 2026). [34]

Longpre et al. (2024), "Consent in Crisis: The Rapid Decline of the AI Data Commons", https://arxiv.org/abs/2407.14933. [38]

McKenzie, T. et al. (2024), "Inverse Scaling: When Bigger Isn't Better", *Transactions on Machine Learning Research*, Vol. 10, https://arxiv.org/abs/2306.09479. [8]

MIC and METI (2026), *AI governance guidance*. [11]

OAIC (2024), *Guidance on privacy and developing and training generative AI models*. [19]

OECD (2026), "The agentic AI landscape and its conceptual foundations", *OECD Artificial Intelligence Papers*, No. 56, OECD Publishing, Paris, https://doi.org/10.1787/396cf758-en. [30]

OECD (2025), *Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence*, OECD Publishing, Paris, https://doi.org/10.1787/23a70dca-en. [6]

OECD (2025), *GPAI workshop on "User control and privacy rights in relation to AI systems"*. [3]

OECD (2025), *Hiroshima AI Process Reporting Framework*, https://transparency.oecd.ai/reports (accessed on 3 February 2026). [21]

OECD (2025), "Intellectual property issues in artificial intelligence trained on scraped data", [33]

OECD Artificial Intelligence Papers, No. 33, OECD Publishing, Paris, https://doi.org/10.1787/d5241a23-en.

OECD (2025), "Mapping relevant data collection mechanisms for AI training", *OECD Artificial Intelligence Papers*, No. 48, OECD Publishing, Paris, https://doi.org/10.1787/3264cd4c-en. [9]

OECD (2025), "Sharing trustworthy AI models with privacy-enhancing technologies", *OECD Artificial Intelligence Papers*, No. 38, OECD Publishing, Paris, https://doi.org/10.1787/a266160b-en. [14]

OECD (2024), "AI, data governance and privacy: Synergies and areas of international co-operation", *OECD Artificial Intelligence Papers*, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/2476b1a4-en. [4]

OECD (2023), "Emerging privacy-enhancing technologies: Current regulatory and policy approaches", *OECD Digital Economy Papers*, No. 351, OECD Publishing, Paris, https://doi.org/10.1787/bf121be4-en. [24]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [2]

OECD (2021), *Recommendation of the Council on Enhancing Access to and Sharing of Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463. [5]

OPC (2023), *Principles for responsible, trustworthy and privacy-protective generative AI technologies*, https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/ (accessed on 3 February 2026). [17]

PDPC (2026), *Proposed Advisory Guidelines on Use of Personal Data in Generative AI*, https://www.pdpc.gov.sg/organisations/regulations-decisions/regulatory-guidance/public-consultation-on-the-proposed-advisory-guidelines-on-use-of-personal-data-in-generative-ai (accessed on 30 June 2026). [20]

PDPC (2025), *Privacy Enhancing Technologies: Adoption Guide*. [25]

PDPC (2024), *Advisory guidelines on use of personal data in AI recommendation and decision systems*, https://www.pdpc.gov.sg/organisations/regulations-decisions/regulatory-guidance/advisory-guidelines-on-use-of-personal-data-in-ai-recommendation-and-decision-systems (accessed on 30 June 2026). [22]

PIPC (2026), *Personal Information Processing Guide for the Development and Use of Generative Artificial Intelligence*, https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=11410 (accessed on 6 March 2026). [23]

UK Government (2026), *Guidelines and best practices for making government datasets ready for AI*. [10]

Villalobos et al. (2024), "Position: Will we run out of data? Limits of LLM scaling based on human-generated data", *ICML'24: Proceedings of the 41st International Conference on Machine Learning*, pp. 49523 - 49544, https://doi.org/10.5555/3692070.3694094. [7]

Zhang et al. (2024), ""It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents", *In Proceedings of the 2024* [16]

CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA 156, pp. 1-26, https://doi.org/10.1145/3613904.3642385.

Contact

Sergi GÁLVEZ DURAN and Christian REIMSBACH-KOUNATZE, dataandprivacy@oecd.org

This work is issued under the responsibility of the Secretary-General of the OECD, and does not necessarily reflect the official views of OECD Member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

© OECD 2026

Attribution 4.0 International (CC BY 4.0)

This work is made available under the Creative Commons Attribution 4.0 International licence. By using this work, you accept to be bound by the terms of this licence (https://creativecommons.org/licenses/by/4.0/).

Attribution – you must cite the work.

Translations – you must cite the original work, identify changes to the original and add the following text: *In the event of any discrepancy between the original work and the translation, only the text of original work should be considered valid.*

Adaptations – you must cite the original work and add the following text: *This is an adaptation of an original work by the OECD. The opinions expressed and arguments employed in this adaptation should not be reported as representing the official views of the OECD or of its Member countries.*

Third-party material – the licence does not apply to third-party material in the work. If using such material, you are responsible for obtaining permission from the third party and for any claims of infringement.

You must not use the OECD logo, visual identity or cover image without express permission or suggest the OECD endorses your use of the work.

Any dispute arising under this licence shall be settled by arbitration in accordance with the Permanent Court of Arbitration (PCA) Arbitration Rules 2012. The seat of arbitration shall be Paris (France). The number of arbitrators shall be one.

[1] While the taxonomy focuses on data collection sources, it is important to note that the ways in which data are made available can vary significantly across these mechanisms. In some cases, data are acquired outright – for example, through commercial licensing agreements, where the licensee may gain ownership or broad reuse rights. In other cases, data are only accessible under defined conditions, such as within secure processing environments, where it can be analysed but not downloaded or retained.

[2] The global chatbot market size was estimated at USD 7,763.0 million in 2024 and is projected to reach USD 27,297.2 million by 2030 (Grand View Horizon, 2026[44]).

[3] As defined in (OECD, 2026[30]), "The agentic AI landscape and its conceptual foundations", OECD Artificial Intelligence Papers, No. 56, OECD Publishing, Paris, AI agents are systems that can perceive and act upon their environment with a degree of autonomy, using tools as needed to achieve specific goals and adapt to changing inputs and contexts. Agentic AI generally refers to systems composed of multiple co-ordinated AI agents that can break down tasks, collaborate, and pursue complex objectives autonomously over extended periods.

[4] Robots.txt is the filename used for implementing the Robots Exclusion Protocol, a standard used by websites to indicate to visiting web crawlers and other web robots which portions of the website they are allowed to visit.