OECD publishing

AI, DATA GOVERNANCE AND PRIVACY

SYNERGIES AND AREAS OF INTERNATIONAL CO-OPERATION

OECD ARTIFICIAL INTELLIGENCE PAPERS

June 2024 **No. 22**

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

The report “AI, data governance, and privacy: Synergies and areas of international co-operation” explores the intersection of AI and privacy and ways in which relevant policy communities can work together to address related risks, especially with the rise of generative AI. It highlights key findings and recommendations to strengthen synergies and areas of international co-operation on AI, data governance and privacy.

This paper was approved and declassified by written procedure by the OECD Digital Policy Committee (DPC) on 20 June 2024 and prepared for publication by the OECD Secretariat. The paper is informed by the contributions of the OECD.AI Expert Group, on AI, Data and Privacy (hereafter the “Expert Group”) of the OECD Network of Experts on AI. It was prepared under the aegis of the OECD Working Party on Artificial Intelligence Governance (AIGO) and the OECD Working Party on Data Governance and Privacy, both working parties of the OECD Digital Policy Committee (DPC).

At the time of publishing, the Expert Group was co-chaired by Reuven Eidelman (Israeli Privacy Protection Authority), Denise Wong (Singapore Infocomm Media Development Authority (IMDA)), and Clara Neppel (IEEE European Business Operations). The Expert Group also benefitted from input and guidance from a Steering Group comprised of Yordanka Ivanova (European Commission), Kari Laumann (Norwegian Data Protection Authority), Winston Maxwell (Télécom Paris - Institut Polytechnique de Paris), and Marc Rotenberg (Center for AI and Digital Policy).

The report development and drafting were led by members of the OECD Secretariat, in collaboration with Winston Maxwell (Télécom Paris - Institut Polytechnique de Paris) who was a major contributor to the report: Sarah Bérubé, Celine Caira, and Yuki Yokomori from the OECD AI Unit, Sergi Galvez Duran, Andras Molnar and Limor Shmerling-Magazanik from the OECD Data Governance and Privacy Unit. Clarisse Girot and Karine Perset, Heads of the Data Governance and Privacy Unit and the AI Unit respectively, recognised the value of the two working parties and the associated policy communities working together and provided resources, input and oversight. Gallia Daor, Digital Economy Policy Division, and Audrey Plonk, Deputy Director of Science, Technology and Innovation, provided advice and oversight.

The authors gratefully acknowledge the contributions made by individuals and institutions that took the time to participate in presentations to the Expert Group.

Finally, the authors thank Andreia Furtado, Marion Barberis, and Shellie Phillips for administrative and communications support, the overall quality of the report benefited from their engagement.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CDEP/AIGO/DGP(2023)1/FINAL

Table of contents

TABLES

BOXES

Acronyms and abbreviations

AI	Artificial intelligence
DPA	Data Protection Authority
DPC	Digital Policy Committee
GPU	Graphics processing unit
HPC	High-performance computing
ICT	Information and communication technology
IGO	Intergovernmental organisation
ML	Machine learning
NGO	Non-governmental organisation
NLP	Natural language processing
OECD	Organisation for Economic Co-operation and Development
ONE AI	OECD AI Network of Experts
PEA	Privacy Enforcement Authority
R&D	Research and development
SDG	Sustainable Development Goals
SME	Small and medium-sized enterprise
VC	Venture Capital
WPAIGO	Working Party on Artificial Intelligence Governance
WPDGP	Working Party on Data Governance and Privacy

Abstract

Recent AI technological advances, particularly the rise of generative AI, have raised many data governance and privacy questions. However, AI and privacy policy communities often address these issues independently, with approaches that vary between jurisdictions and legal systems. These silos can generate misunderstandings, add complexities in regulatory compliance and enforcement, and prevent capitalising on commonalities between national frameworks. This report focuses on the privacy risks and opportunities stemming from recent AI developments. It maps the principles set in the OECD Privacy Guidelines to the OECD AI Principles, takes stock of national and regional initiatives, and suggests potential areas for collaboration. The report supports the implementation of the OECD Privacy Guidelines alongside the OECD AI Principles. By advocating for international co-operation, the report aims to guide the development of AI systems that respect and support privacy.

Abrégé

Les récentes avancées technologiques en matière d'IA, en particulier l'essor de l'IA générative, ont soulevé de nombreuses questions concernant la gouvernance des données et la protection de la vie privée. Cependant, les communautés de l'IA et de la politique de protection de la vie privée abordent souvent ces enjeux de manière indépendante, avec des approches qui varient d'une juridiction à l'autre et d'un système juridique à l'autre. Ces cloisonnements peuvent générer des malentendus, ajouter des complexités au respect et à l'application des réglementations et empêcher de capitaliser sur les points communs entre les cadres nationaux. Ce rapport se concentre sur les risques et les opportunités en matière de protection de la vie privée découlant des récents développements de l'IA. Il compare les Lignes directrices de l'OCDE en matière de protection de la vie privée avec les Principes de l'OCDE en matière d'IA, fait le point sur les initiatives nationales et régionales et suggère des domaines potentiels de collaboration. Le rapport soutient la mise en œuvre des Lignes directrices de l'OCDE relatives à la protection de la vie privée parallèlement aux Principes de l'OCDE relatifs à l'IA. En prônant la coopération internationale, le rapport vise à guider le développement de systèmes d'IA qui respectent et soutiennent la protection de la vie privée.

Executive summary

Recent AI technological advances—particularly the rise of generative AI— raise both opportunities and risks related to data protection and privacy. As a general-purpose technology, AI is wide-reaching and rapidly permeating products, sectors, and business models across the globe. Recent progress in generative AI owes its progress in large part to the availability and use of vast training data stored worldwide. Like the data, actors involved in the AI lifecycle are distributed across jurisdictions, underscoring the need for global synchronisation, clear guidance and cooperative efforts to address the challenges posed by AI's impact on privacy.

However, the AI and privacy policy communities currently tend to address challenges separately, without much co-operation, such that their approaches vary across jurisdictions and legal systems. For instance, the practice of scraping personal data to train generative AI raises significant privacy questions and is attracting increasing regulatory attention as a consequence. However, discussions on practical solutions to align data scraping practices with Privacy Guidelines have been limited. Likewise, the practical implementation of individual data protection and privacy rights in the development of Generative AI is not yet the subject of collective in-depth reflection. As more countries move to regulate AI, lack of co-operation between these communities could result in misunderstandings as to the actual reach of data protection and privacy laws, as well as in conflicting and/or duplicative requirements that could lead to additional complexity in regulatory compliance and enforcement. As both communities consider possible responses to the opportunities and risks of AI, they could benefit from each other's knowledge, experience, and priorities through greater collaboration, aligning policy responses and improving complementarity and consistency between AI policy frameworks on the one hand, and data protection and privacy frameworks on the other.

With their differences in history, profiles and approaches, the AI and privacy policy communities have lessons to learn from each other. In recent years, the AI community, including AI researchers and developers, from academia, civil society, and the public and private sectors, has formed dynamic and strong networks. Many in the AI community have taken an innovation-driven approach, while the privacy community has been generally adopting a more cautious approach marked by decades of implementation of long-standing privacy and data protection laws. The privacy community is also often characterised as more established because of long-standing privacy and data protection laws and has evolved over time to include a diverse array of stakeholders such as regulators, privacy and data protection officers, technologists, lawyers, public policy professionals, civil society groups, and regulatory technology providers, among others. This community is focused on establishing privacy safeguards and mitigating risks assessed within often sophisticated and firmly established regulatory frameworks. Despite these differences, synergies exist, and co-operation is essential.

This report identifies areas that would benefit from further synergy and complementarity, including key terminological differences between the two policy communities. It maps existing privacy and data protection considerations to the AI values-based principles set in the OECD's 2019 Recommendation on AI to identify relevant areas for closer coordination. This mapping illustrates the different interpretations of

the privacy and AI communities around key concepts – including fairness, transparency and explainability. Understanding these differences is essential for building sustainable co-operation actions.

Actors in both the AI and privacy communities have implemented measures at the national, regional, and global levels to tackle opportunities and risks posed by AI. The report provides a snapshot of national and regional developments on AI and privacy, including guidance provided by privacy regulators on the application of privacy laws to AI and related enforcement actions, specifically regarding generative AI. It finds that while many actions have been taken, including policy initiatives and enforcement actions by Privacy Enforcement Authorities, they could benefit from further coordination as AI-specific laws emerge worldwide.

With its international reach and substantive expertise in both AI and data protection and privacy, the OECD appears as a key forum to strengthen synergies and areas of international co-operation in this area. It can draw on well-established policy work in both areas, including the leading principles included in the 1980 OECD Privacy Guidelines, updated in 2013, and the 2019 OECD Recommendation on AI, updated in 2024. Moreover, in 2024, the OECD has established a unique Expert Group on AI, Data, and Privacy, which convenes leading voices in both communities to explore key questions and policy solutions at the intersection of AI and data protection and privacy.

Despite the challenges, both this policy work and the ongoing activities within this expert group showcase that broad and lasting co-operation, as well as mutual understanding, are achievable. To provide a common reference framework for these co-operation opportunities and highlight the OECD's distinctive role, the report aligns the OECD AI Principles—the first intergovernmental standard on AI—with the well-established OECD Privacy Guidelines, which serve as the foundation for data protection laws globally.

The report assesses national and regional initiatives related to AI and privacy and identifies areas for collaboration, such as in the area of Privacy Enhancing Technologies (PETs), that can help address privacy concerns, particularly regarding the "explainability" of AI algorithms. The joint expert group on AI, data, and privacy will play a crucial role, in more precisely articulating the concrete opportunities for innovative, technological and regulatory developments of AI that respect privacy and personal data protection rules.

Résumé

Les récentes avancées technologiques en matière d'IA – en particulier l'essor de l'IA générative – soulèvent à la fois des opportunités et des risques liés à la protection des données et de la vie privée. En tant que technologie polyvalente, l'IA a une grande portée et pénètre rapidement les produits, les secteurs et les modèles d'entreprise dans le monde entier. Les progrès récents de l'IA générative sont dus en grande partie à la disponibilité et à l'utilisation de vastes données d'entraînement stockées dans le monde entier. Tout comme les données, les acteurs impliqués dans le cycle de vie de l'IA sont répartis entre différentes juridictions, ce qui souligne la nécessité d'une synchronisation mondiale, d'orientations claires et d'efforts de coopération pour relever les défis posés par l'impact de l'IA sur la vie privée.

Toutefois, les communautés de l'IA et de la politique de protection de la vie privée tendent actuellement à relever les défis séparément, sans grande coopération, de sorte que leurs approches varient d'une juridiction à l'autre et d'un système juridique à l'autre. Par exemple, la pratique consistant à gratter (« scraping ») des données personnelles pour entraîner l'IA générative soulève des questions importantes en matière de protection de la vie privée et suscite de ce fait une attention croissante sur le plan réglementaire. Cependant, les discussions sur les solutions pratiques permettant d'aligner les pratiques de récupération de données sur les principes de protection de la vie privée ont été limitées. De même, la mise en œuvre pratique de la protection des données individuelles et des droits à la vie privée dans le développement de l'IA générative ne fait pas encore l'objet d'une réflexion collective approfondie. Alors que de plus en plus de pays s'apprêtent à réglementer l'IA, le manque de coopération entre ces communautés pourrait entraîner des malentendus quant à la portée réelle des lois sur la protection des données et de la vie privée, ainsi que des exigences contradictoires et/ou redondantes susceptibles d'accroître la complexité du respect et de l'application de la réglementation. Alors que les deux communautés envisagent des réponses possibles aux opportunités et aux risques de l'IA, elles pourraient bénéficier de leurs connaissances, de leur expérience et de leurs priorités respectives grâce à une plus grande collaboration, en alignant les réponses politiques et en améliorant la complémentarité et la cohérence entre les cadres politiques de l'IA, d'une part, et les cadres de protection des données et de la vie privée, d'autre part.

Avec leurs différences d'histoire, de profils et d'approches, les communautés de l'IA et de la politique de protection de la vie privée ont des leçons à tirer l'une de l'autre. Ces dernières années, la communauté de l'IA, qui comprend des chercheurs et des développeurs dans le domaine de l'IA, issus du monde universitaire, de la société civile et des secteurs public et privé, a formé des réseaux dynamiques et solides. De nombreux membres de la communauté de l'IA ont adopté une approche axée sur l'innovation, tandis que la communauté de la protection de la vie privée est souvent considérée comme plus établie en raison de l'existence de lois de longue date sur la protection de la vie privée et des données. La communauté de la protection de la vie privée a évolué au fil du temps pour inclure un éventail diversifié de parties prenantes telles que des régulateurs, des responsables de la protection de la vie privée et des données, des technologues, des juristes, des professionnels des politiques publiques, des groupes de la société civile et des fournisseurs de technologies de régulation, entre autres. Cette communauté se concentre sur la mise en place de garanties en matière de protection de la vie privée et sur l'atténuation

des risques évalués dans des cadres réglementaires souvent sophistiqués et solidement établis, adoptant généralement une approche plus prudente que celle qui tend à caractériser la communauté de l'IA. Malgré ces différences, des synergies existent et la coopération est essentielle.

Ce rapport identifie les domaines qui bénéficieraient d'une synergie et d'une complémentarité accrues, y compris les principales différences terminologiques entre les deux communautés politiques. Il compare les considérations existantes en matière de protection de la vie privée et des données avec les principes fondés sur les valeurs de l'IA énoncés dans la Recommandation de 2019 de l'OCDE sur l'IA, afin d'identifier les domaines pertinents pour une coordination plus étroite. Cette cartographie illustre les différentes interprétations des communautés de la protection de la vie privée et de l'IA autour de concepts clés – notamment l'équité, la transparence et l'explicabilité. Il est essentiel de comprendre ces différences pour mettre en place des actions de coopération durables.

Les acteurs des communautés de l'IA et de la protection de la vie privée ont mis en œuvre des mesures aux niveaux national, régional et mondial pour faire face aux opportunités et aux risques posés par l'IA. Le rapport donne un aperçu des évolutions nationales et régionales en matière d'IA et de protection de la vie privée, y compris les orientations fournies par les autorités de régulation de la protection de la vie privée sur l'application des lois sur la protection de la vie privée à l'IA et les mesures d'exécution connexes, en particulier en ce qui concerne l'IA générative. Il constate que si de nombreuses mesures ont été prises, notamment des initiatives politiques et des mesures d'application par les autorités chargées de l'application des lois sur la protection de la vie privée, elles pourraient bénéficier d'une plus grande coordination à mesure que des lois spécifiques à l'IA voient le jour dans le monde entier.

Grâce à sa portée internationale et à son expertise dans les domaines de l'IA, de la protection des données et de la vie privée, l'OCDE apparaît comme un forum essentiel pour renforcer les synergies et les domaines de coopération internationale dans ce domaine. Elle peut s'appuyer sur des travaux politiques bien établis dans les deux domaines, notamment les principes fondamentaux inclus dans les Lignes directrices de l'OCDE relatives à la protection de la vie privée de 1980, mises à jour en 2013, et la Recommandation de l'OCDE sur l'IA de 2019, mise à jour en 2024. En outre, en 2024, l'OCDE a créé un groupe d'experts unique sur l'IA, les données et la protection de la vie privée, qui réunit des personnalités de premier plan des deux communautés afin d'examiner les questions clés et les solutions politiques à l'intersection de l'IA et de la protection des données et de la vie privée.

Malgré les défis, le travail de politiques publiques et les activités en cours au sein de ce groupe d'experts montrent qu'une coopération large et durable, ainsi qu'une compréhension mutuelle, sont réalisables. Afin de fournir un cadre de référence commun pour ces opportunités de coopération et de souligner le rôle distinctif de l'OCDE, le rapport aligne les Principes de l'OCDE relatifs à l'IA – la première norme intergouvernementale sur l'IA – sur les Lignes directrices de l'OCDE relatives à la protection de la vie privée, qui servent de fondement aux lois sur la protection des données à l'échelle mondiale.

Le rapport évalue les initiatives nationales et régionales liées à l'IA et à la protection de la vie privée et identifie les domaines de collaboration, tels que les technologies d'amélioration de la protection de la vie privée (PET), qui peuvent contribuer à répondre aux préoccupations en matière de protection de la vie privée, en particulier en ce qui concerne l'« explicabilité » des algorithmes d'IA. Le groupe d'experts conjoint sur l'IA, les données et la vie privée jouera un rôle crucial en définissant plus précisément les possibilités concrètes de développement innovant, technologique et réglementaire de l'IA dans le respect de la vie privée et des règles de protection des données à caractère personnel.

Introduction

Strengthening synergies between the AI and privacy communities

Recent AI advancements, including the rise of generative AI in late 2022, have raised data governance and privacy challenges. Difficult questions have come to the fore around the use of input and output data, data quality and data availability for training AI models. Namely, how to protect the rights and interests of all parties involved, including individuals to whom the data collected, used, and produced by these models and systems relate.

In contrast to previous AI systems, recent advances in neural networks and deep learning have resulted in larger, more advanced and more compute-intensive AI models and systems. In 2017, a group of researchers introduced a type of neural network architecture called "transformers" – a key conceptual breakthrough underpinning major progress in AI language models and generative AI. These advances centre on "foundation models" – models trained on large amounts of data that can be adapted to a wide range of downstream tasks such as OpenAI's Generative Pretrained Transformers (GPT) series. Advances in AI computing infrastructure – such as graphics processing units (GPUs) – and the availability and quality of data, have also been fundamental in fuelling technological leaps in machine learning, as they form the basic AI production function: algorithms, data and computing resources (OECD, 2024[1]).

With the rise of new machine learning techniques, and generative AI applications in particular, calls have become pressing to consider the privacy implications related to the training and use of AI systems, and for the different policy communities in this area – including policymakers, researchers, civil society, industry, and oversight and enforcement agencies – to help address such concerns by cross-fertilising their respective efforts. Neural networks have often been referred to as a "black box". The term "black box" reflects the considerable challenge in understanding how AI systems make decisions, a challenge that is especially apparent in neural networks-based methods. The OECD is helping to build and strengthen synergies between the AI and privacy communities, drawing on well-established policy work in both areas, and on the leading principles included in the 1980 OECD Privacy Guidelines, updated in 2013, and the 2019 OECD Recommendation on AI (hereafter "OECD Recommendation on AI"), updated in 2024.

This analysis concludes that, despite challenges, AI's innovative, technological and regulatory developments are mainly compatible with, and can even reinforce, privacy and personal data protection rules. By identifying risks and opportunities, mapping existing OECD Privacy Guidelines to AI Principles, taking stock of national and regional initiatives, and providing key policy considerations for the way forward, it advances the OECD's mission to help implement the OECD AI Principles, the world's first intergovernmental standard on AI, and the well-established OECD Privacy Guidelines, a flagship legal instrument which forms the bedrock of data protection laws globally. Such efforts help to enable international co-operation, at the OECD and beyond, to foster a shared understanding to help chart the course for successful implementation of AI and privacy rules around the globe.

AI and privacy: a dynamic policy landscape

In response to the recent rise of advanced machine learning systems and generative AI, many in both the AI and the privacy and data protection communities are posing questions at the intersection of AI, data governance, and privacy, and organising policy responses. Policy actions and rules around AI systems applied in high-risk areas have also emerged in a growing number of jurisdictions. In likely the best-known example, the European Union's Artificial Intelligence Act (the 'EU AI Act') outlines a risk-based regulatory approach to the use of AI systems, including areas of high-risk, for example threats to European values such as privacy and data protection (European Parliament, 2024[2]). G7 Digital and Technology Ministers have also put AI high on their agendas through the G7 Hiroshima Process on Generative AI (OECD, 2023[3]), emphasising the need to protect human rights, including the right to privacy. Individual countries have also taken actions. For example, the 2023 United States Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, directs public sector entities to establish new standards for AI safety and security, privacy protections, equity and civil rights, consumers' and workers' rights, and innovation and competition (The White House, 2023[4]).

In parallel to the establishment of AI-specific laws, regulations and standards, regulators in a growing number of jurisdictions have started applying existing privacy and data protection laws (hereinafter 'privacy laws' for short) to address the privacy issues generated by the processing of personal data used to train AI systems, such as through enforcement actions or the adoption of ad hoc guidance (see Section 3). Several Privacy Enforcement Authorities (PEAs)[1] announced the launch of AI action plans, including the establishment of dedicated AI units. These national initiatives are reinforced by initiatives from the international community of privacy authorities, for example the public statement issued by the G7 Data Protection and Privacy Authorities (DPAs) Roundtable in June 2023 and the "Resolution on Generative Artificial Intelligence Systems" adopted at the 45th meeting of the Global Privacy Assembly, the premier global network of privacy regulators, on 20 October 2023.

With variations between jurisdictions and legal systems, however, the AI and privacy policy communities are still largely responding to AI and privacy challenges independent of each other. As more countries move to regulate AI, this could result in misunderstandings on the actual reach of data protection and privacy laws, and in conflicting and/or duplicative requirements that can lead to additional complexity in compliance and enforcement of responsible policies and regulations. As both communities consider possible responses to AI opportunities and risks, they could benefit from each other's knowledge, experience, and priorities through greater collaboration, aligning policy responses and improving complementarity and consistency between AI policy frameworks on the one hand, and data protection and privacy frameworks on the other.

The existence of parallel work streams involving two different policy communities is not unusual nor is it necessarily problematic. Each community brings a unique perspective that can lead to a richer policy debate and approach to solutions. Rapid advances in AI as a technology and its diffusion across sectors are, however, putting pressure on different policy communities to devise solutions quickly, and the existence of silos raises the risk of inconsistent policy responses and even misunderstandings due to differences in terminology and approaches. Achieving consensus on language is often a pre-requisite for effective co-operation.

In this vein, the OECD has played an important role promoting standardisation of key terminology in the AI and privacy space. Namely, the 2019 OECD Recommendation on AI includes a widely cited definition of an AI system which was revised in late 2023 to ensure it reflects and addresses important technology and policy developments, notably with respect to generative AI, and heightened concerns around safety, information integrity, and environmental sustainability. The Recommendation has influenced AI policy and legal frameworks around the world, including in the EU AI Act, the Council of Europe, and in standards bodies like the US National Institute of Standards and Technology (NIST). The EU-US Trade and

Technology Council (TTC) is also actively collaborating and released an initial draft of a common EU-US Terminology and Taxonomy for Artificial Intelligence on 31 May 2023, which includes terms relevant to both the AI and privacy communities (European Commission and US TTC, 2023[5]).

The OECD Working Party on Artificial Intelligence Governance (WPAIGO) and the OECD Working Party on Data Governance and Privacy (WPDGP) are well positioned to support existing international co-operation efforts on these issues. As part of the OECD.AI Network of Experts (ONE AI) the OECD.AI Expert Group on AI, Data, and Privacy (hereafter 'Expert Group') established in 2024, also helps to bring both communities together to promote synergies and complementariness. Such OECD Working Party delegates and members of the Expert Group have contributed analysis and insights to this report and work more broadly.

The OECD privacy and data protection community is well-established, with a robust "toolbox" relevant to risks raised by AI to individual rights and freedoms. This toolbox comprises various legal instruments, notably the OECD Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (hereafter "OECD Privacy Guidelines"), adopted in 1980 and revised in 2013 [OECD/LEGAL/0188] and the OECD Recommendation on Enhancing Access to and Sharing of Data [OECD/LEGAL/0463], as well as relevant materials and expertise in implementing fundamental concepts and compliance mechanisms in different contexts. Elements of this toolbox are already being used in several jurisdictions to inform frameworks for the responsible use of AI.

While frameworks related to trustworthy AI are comparatively more recent, they have garnered significant attention globally. In efforts to keep pace with the speed of technological advancements, the AI policy community is rapidly developing and implementing frameworks including on AI risk management and accountability, mitigating bias, promoting explainability of AI system outputs, and improving robustness, among others, along the AI system lifecycle. For example, the OECD's 2019 Recommendation on AI [OECD/LEGAL/0449], updated in 2024, forms a foundational set of principles that has guided the development of AI laws, regulations and standards globally.

The rapidly evolving AI, data governance, and privacy policy landscape calls for greater collaboration and information sharing across these policy communities. Co-operation strives to support efforts not only to address privacy risks in AI, but also to optimise and enhance the collective benefits of AI to society, including both unleashing AI's innovative potential, while also protecting privacy and personal data. International co-operation on AI and privacy requires ensuring the long-term interoperability of legal, technical, and operational frameworks applying to AI and privacy. This will allow policy- and decision-makers to leverage commonalities, complementarities, and elements of convergence in their respective policy frameworks, or, conversely, to identify the stumbling blocks that could hinder the development of common positions or co-operation. These co-operation efforts could help assess whether the AI and Privacy Recommendations need to be updated to reflect the synergies between the AI and privacy communities.

AI and the OECD Privacy Guidelines

On the privacy side, the OECD Privacy Guidelines, were adopted in 1980 and revised in 2013 [OECD/LEGAL/0188]. They are the cornerstone of the OECD's work on privacy and are recognised as the global minimum standard for privacy and data protection.

The OECD Privacy Guidelines are complemented by other flagship OECD legal instruments relevant to co-operation on privacy and AI issues, including the Recommendation on Cross-border Co-operation on Enforcement of Laws Protecting Privacy (OECD, 2007[6]), the Recommendation on Enhancing Access to and Sharing of Data (OECD, 2021[7]), and subsequent implementation guidance and work carried out under the auspices of the WPDGP. The Declaration on Government Access to Personal Data Held by

Private Sector Entities [OECD/LEGAL/0487] is also relevant when accessing personal data stored in AI systems for national security and law enforcement purposes. In recent years, the WPDGP has produced analyses relevant to privacy and AI, in particular on Privacy Enhancing Technologies (PETs) (OECD, 2023[8]).

International co-operation is a core principle of the OECD Privacy Guidelines (Part Six) and a growing area of work in the WPDGP's agenda, including on the need to clarify the intersection of baseline privacy frameworks with sectoral or other cross-cutting frameworks that include AI and new technologies.

The OECD Privacy Guidelines are technology neutral and do not explicitly cover the privacy challenges posed by AI nor of other specific digital technologies. At the same time, the need to address the potential for bias and other harmful consequences from personal data processing without hindering innovation and preventing the beneficial uses of these technologies was highlighted in the 2021 review of the Recommendation (OECD, 2021[9]).

One of the dominant themes is the importance of "explainability" of AI algorithms to ensure accuracy, fairness, and accountability. Experts also noted that AI increases demand for large data sets, which are critical to build AI systems that generate more accurate outputs, but also increase privacy-related risks. Furthermore, experts highlighted that most AI Principles refer to privacy in general terms but do not establish an explicit connection between the capabilities of AI and the nature of AI-specific privacy challenges, with the possible effect of shifting the focus away from privacy when it comes to AI. Their overall suggestion was that additional guidance may be helpful to ensure that current AI Principles sufficiently address privacy-related concerns.

Adherents to the OECD Privacy Guidelines agreed with these points at the time. They noted that the technology-neutral language of the OECD Privacy Guidelines was key to their adaptability and decided not to amend the basic principles to account for AI (OECD, 2021[9]). Rather, Adherents decided that these important matters related specifically to AI could be addressed in mechanisms and guidance related to the OECD Recommendation on AI, which had just been adopted in 2019. The analysis undertaken in this report, and the work of the Expert Group, also demonstrate ways fulfilling the request for further collaboration with AI communities.

Privacy and the OECD Recommendation on AI

Since 2016, the OECD has undertaken significant work on AI policy and governance through its Working Party on AI Governance (AIGO) (OECD AI, 2023[10]), including the adoption of the OECD Recommendation of the Council on Artificial Intelligence in May 2019, updated in 2024. Comprising five principles applicable to all stakeholders and five recommendations to governments, the OECD AI Principles provide guidance on how governments and other actors can shape a human-centric approach to trustworthy AI. With 46 adherents, the AI Principles emphasise international collaboration, including in the areas of privacy and accountability.

The global significance of the OECD AI Principles cannot be overstated. Since their adoption, countries worldwide have taken actions to codify the Principles through national AI strategies, and soft and hard laws. These include several AI initiatives, including the establishment of AI Offices or AI Commissioners to guide implementation of national or regional laws, regulations and standards, for example, in the EU AI Act, and Canada's proposed Artificial Intelligence and Data Act (AIDA). The United States Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, also emphasizes the role of the Federal Trade Commission to ensure fair competition in the AI marketplace and to ensure that consumers and workers are protected from AI harms. Some jurisdictions are also exploring the role of existing national PEAs in implementing AI-related law and regulation.

Exploring the role of existing data protection actors in the implementation of AI frameworks may not come as a surprise, considering values-based principles such as the OECD AI Principles complement existing OECD standards in areas such as privacy, with privacy and data protection being core components of the Recommendation. Three principles in particular mention data protection and privacy explicitly in the OECD AI Principles (OECD, 2019[11]):

- **Principle 1.2 Respect for the rule of law, human rights, and democratic values, including fairness and privacy,** throughout the AI system lifecycle.
- **Principle 1.4. Robustness, security and safety** calls for AI actors to, among other things, ensure the traceability of AI systems including in relation to datasets, processes and decisions made during the AI system lifecycle, and that AI actors should apply a systematic risk management approach to AI system lifecycle phases to address risks such as privacy, digital security, safety and bias.
- **Principle 2.1. Investing in AI research and development** calls for, among other things, governments to consider public investment and encourage private investment in open datasets that are representative and respect privacy and data protection.

Privacy is also referenced in many tools included in the work of AIGO and ONE AI. For example, the OECD Framework for the Classification of AI Systems has been applied to contexts where privacy and data protection is paramount, including in evaluating medical technology applications in the United Kingdom and in Australia.

Key definitions of terms used throughout the report

AI and privacy communities

This report refers to “AI and privacy communities”, which for the purposes of this analysis is defined as including policymakers, enforcement authorities, as well as experts in academia, civil society, and private sector professionals.

The “AI community”, including AI researchers and developers, and members from academia, civil society, and the public and private sectors, has grown in recent years into a dynamic and strong network focused both on advancing highly technical aspects of AI as well as global AI governance to promote its responsible use. This group is subject to established or emerging rules including AI laws, regulatory frameworks, and standards, which in many parts of the world are still evolving. Many in the AI community have taken an innovation-driven approach, including the exploration of largely uncharted territories as technological leaps and new applications are discovered.

The “privacy community”, in contrast, may be characterised as more established, although its profile has dramatically evolved over time to include a very diverse array of stakeholders including regulators, privacy and data protection officers, technologists, lawyers, public policy professionals, civil society groups, and regulatory technology providers, among others. This broad global community has been largely shaped by the growing number of regulatory frameworks that have developed over decades around the globe. Because it operates in this mature context, the privacy community adopts approaches that can generally be described as more cautious than the innovation-driven approaches which tend to characterise the AI community.

The differing nature of these communities’ approaches to AI and privacy issues has implications. While the AI community may benefit from its agility and innovative spirit, it might lack experience with and in-depth understanding of the regulatory implications of technology advancements, regulatory implementation and enforcement of AI-specific rules. Meanwhile, the privacy community, with its depth of regulatory experience, may lack the technical knowledge necessary to fully comprehend how personal

data is used in designing, developing, and deploying AI systems. This technical gap could lead to an overly conservative approach, potentially hindering innovation due to concerns over the privacy risks associated with using personal data to train and test AI systems.

Building bridges between these communities would not only facilitate compliance with emerging AI laws, but also ensure that AI development continues to thrive within existing data protection and privacy frameworks. A collaborative approach is essential for formulating regulations that protect societal values without impeding technological progress.

AI actors

According to the OECD AI Principles, the term “AI actors” refers to those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI. An AI system lifecycle typically involves several phases that include to: plan and design; collect and process data; build model(s) and/or adapt existing model(s) to specific tasks; test, evaluate, verify and validate; make available for use/deploy; operate and monitor; and retire/decommission. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

Generative AI

Generative artificial intelligence (AI) systems create new content (e.g. text, image, audio, or video) in response to prompts, based on the data the models have been trained on. Generative AI is based on machine learning (ML), which has developed gradually since the 1950s. ML models leverage deep neural networks to emulate human intelligence (i.e. by imitating information processing of neurons in the human brain) by being exposed to data (training) and finding patterns that are then used to process previously unseen data. This allows the model to generalise based on probabilistic inference (i.e., informed guesses) rather than causal understanding. Unlike humans, who learn from only a few examples, deep neural networks need hundreds of thousands, millions, or even billions, meaning that machine learning requires vast quantities of data (Lorenz, Perset and Berryhill, 2023[12]).

Privacy and data protection

The terms “privacy” and “data protection” can have different meanings in the AI policy and privacy policy communities, also factoring in that the concept of “data protection”, as a contraction of “the protection of individuals with regard to the processing of their personal data”, is itself already frequently the subject of misinterpretations. Some members of the AI policy community may in particular view “privacy” as relating principally to the risk of re-identification, data leakage or inferences from AI (OECD, 2023[13]), in practice subsuming privacy and data protection, possibly the larger concept of data governance, and the concept of AI safety. As is evident in the work of privacy authorities on AI systems, however, privacy and data protection go beyond issues of safety and security, as explained in an OECD report (OECD, 2023[14]) and in the Resolution of the Global Privacy Assembly (GPA, 2023[15]). Nevertheless, the risk that the AI community treats privacy and data protection as a well-defined “box” to be “ticked” (OECD, 2023[16]), may lead to underestimating the role of privacy in addressing many of the human rights impacts resulting from AI.

Effective coordination requires a common understanding of the terminology in each domain. The basic concepts, tests, and rules that policymakers, and in particular regulators, use in the privacy space provide existing obligations are therefore important to know for the AI policy community. Even more so as this community is itself is moving towards an era where AI regulation is being implemented in practice and may overlap and complement existing privacy and data protection rules.

1 Generative AI: a catalyst for collaboration on AI and privacy

The need for closer coordination between the AI and privacy communities has been identified for some time. But this need has become even more evident, and urgent, with the emergence of generative AI systems, including language models, that generate various forms of content (e.g. text) based on patterns found in vast volumes of training data. While generative AI creates new opportunities across industries and sectors, including in code development, creative industries and arts, education, healthcare, and more (OECD, 2023[13]), this technology also raises new risks, as well as amplifying existing ones, including discrimination, polarisation, opaque decision-making, or potential social control.

The opportunities and risks of generative AI for privacy

The OECD has contributed to analytical work and raising awareness around the opportunities and risks posed by AI systems in relation to privacy and data protection, including those posed by generative AI, through its paper on "Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI" (OECD, 2023[14]) and its "Framework for the Classification of AI systems" (OECD, 2022[17]). Both the AI system lifecycle and the classification of AI systems can provide useful structure for discussion as the nature of privacy challenges will vary based on the phase of the lifecycle and the type of AI system at stake. Some of these opportunities and risks are explored below.

AI training techniques and other tools may bring new opportunities for enhancing privacy (Privacy Enhancing Technologies)

While there are many questions around possible threats to privacy raised by AI, emerging technology applications also bring new opportunities for enhanced privacy protection (OECD, 2024[1]). These reinforce recent and ongoing OECD work on emerging Privacy-Enhancing Technologies (PETs) (OECD, 2023[8]). PETs refer to a range of digital technologies and techniques that enable the collection, processing, analysis, and sharing of information while safeguarding data confidentiality and privacy. Although many emerging PETs are still in their early stages of development, some of them hold immense potential to advance privacy-by-design principles in AI and foster trust, including with regard to data sharing and re-use across organisations.

For example, researchers are developing different encrypted data processing tools which allow data to remain encrypted while in use and thus can help enhance privacy at various stages throughout the AI system lifecycle (OECD, 2023[18]). Such techniques include homomorphic encryption and trusted execution environments (TEEs), where actors along the AI lifecycle are not able to view the underlying data without permission (O'Brien, 2020[19]; Mulligan et al., 2021[20]).

Other techniques allow executing analytical tasks upon data that are not visible or accessible to those executing the tasks (federated and distributed analytics). For instance, federated learning enables developers to train a model using data within their own network, which is then transferred to a central

server to combine the data into an improved model that is shared back with all the users (MIT, 2022[21]). Federated learning solutions are beginning to be implemented in healthcare, showing positive results. That said, ensuring the accessibility of health data for both primary and secondary use purposes remains critical for the development and effective use of AI in healthcare, making it a crucial policy concern.

Other techniques increase privacy protections at both the training and use stage by altering the data, by adding "noise" or by removing identifying details. Among such "data obfuscation techniques'", "differential privacy" algorithms ensure that the output of the AI system minimally changes when a single point of data about an individual is added or retrieved from the training dataset (Harvard, 2024[22]). The technique of synthetic data has also attracted significant interest as a PET approach. Synthetic data are generated via computer simulations, machine-learning algorithms, and statistical or rules-based methods, while preserving the statistical properties of the original dataset.[2] They can be used to train AI when data are scarce or contain confidential or personally identifiable information. These can include datasets on minority languages; training computer vision models to recognise objects that are rarely found in training datasets; or data on different types of possible accidents in autonomous driving systems (OECD, 2023[23]). However, challenges remain. Similar to anonymisation and pseudonymisation, synthetic data can be susceptible to re-identification attacks (Stadler, Oprisanu and Troncoso, 2020[24]), and "[r]e-identification is still possible if records in the source data appear in the synthetic data" (OPC, 2022[25]). Furthermore, some research shows that models trained over a high volume of synthetic data can collapse over time (Shumailov, 2023[26]). In other words, while synthetic data can help fill in some gaps and improve knowledge, it cannot be expected to entirely replace real-world data.

Machine unlearning is another emergent subfield of machine learning that would grant individuals control over their personal data, even after it has been shared. Indeed, recent research has shown that in some cases it may be possible to infer with high accuracy whether an individual's data was used to train a model even if that individual's data has been deleted from a database. Machine unlearning aims to tackle this challenge and would give individuals the possibility to withdraw their consent to the collection and processing of their data and to ask for the data to be deleted, even after they have been shared (Tarun, 2023[27])

In consideration of the many promises which PETs hold for enabling data sharing and the next-generation data economy model, the OECD is considering additional deliverables and further synergies to highlight their potential. A key focus concerns established and emerging AI-related use cases in the public and private sector, including health and finance. Future work will also explore how governments and regulators can best incentivise innovation in and with PETs and discuss how to measure and compare the effectiveness and impact of different techniques.

AI systems, and generative AI in particular, raise privacy concerns

Technical breakthroughs have fueled the development of generative AI systems that are so advanced that users may not be able to distinguish between human and AI-generated content. While such developments are impressive on the technological level, the large amounts of data required to train large AI models, including increasing amounts of personal data acquired through various means, raise serious questions around risks to privacy and data protection.

Generative AI poses privacy risks during both its development and deployment stages. Many developers depend on publicly accessible sources for training data, which often includes data about individuals shared online. However, just because data is accessible does not automatically mean that it is free to be collected and used to train AI models. The collection of personal data for training AI systems, like any data processing activity, is subject to the privacy principles set forth in the OECD Privacy Guidelines and in data protection laws globally. Among others, these principles require that personal data be obtained through lawful and fair means, with the knowledge of the data subject, and that any further uses of the data are not incompatible with the original purposes. While individuals may have shared their data consenting to

another use or uses, these do not necessarily include training AI models (GPA's International Enforcement Cooperation Working Group, 2023[28]).

More recent research shows that generative AI models are actually able to infer personal attributes of the data subject from large collections of unstructured text (e.g. public forum or social network posts) with high accuracy, yet at a low cost (Robin Staab, 2023[29]). This could result in inferences based on gender, race or age data that exacerbate the risk of harmful bias and discrimination.

Moreover, some research pre-dating the prominence of generative AI models already suggested (Ahmed Salem, 2018[30]) that de-identification, which has been used historically to find a balance between using the data in aggregate and protecting people's privacy, does not scale to big data datasets. In certain situations, it is possible to reconstruct and de-anonymise original training data by analysing the behaviour of a model that includes it.

Such concerns are compounded with the inherent lack of transparency in data processing, in possible contradiction with the "Openness Principle" in the OECD Privacy Guidelines and with related information requirements in national laws. Thus, given the capacity of AI models to "memorise" significant volumes of training data, large language models behind text-based generative AI tools pose a particular risk of collection, use and re-use of personal data without the knowledge of the persons concerned (Hannah Brown, 2022[31]).

AI systems, and generative AI specifically, can also appear to be in tension with individuals' rights to access, correct, and where necessary delete their personal data (also known as the "Individual Participation Principle"). Where personal data are used to train machine learning models, their deletion or correction can be complicated, for example because they require additional resources to retrain the model. Furthermore, ensuring these rights in the context of generative AI models might be difficult when training data includes unstructured information curated from the Internet. It might be challenging and resource intensive to identify the data point associated with an individual in an unstructured dataset.

Interactions with users and the feedback loops of autonomous self-learning models may lead to a degradation of the model's accuracy and reliability, introducing the risk of hallucinations" or other forms of misleading content, disinformation, or misinformation. The privacy concern arises from the fact that this new data generated by making inferences can reveal personal information that either has not been disclosed by the individual or is inaccurately attributed to the individual. These various forms of misleading content can also result in security vulnerabilities (Solove, 2024[32]), especially when the AI system is deployed in specific contexts, such as law enforcement, medicine, education, or employment. A question that arises then is whether the privacy rights of individuals are adequately tailored to address these concerns *ex post*. For instance, if a "hallucination" includes inaccurate information including personal data generated by AI, do individuals have a right to have their data corrected and/or deleted? And if the identification and deletion of specific data sets from an AI model is extremely complex, both technically and logistically, to the point of rendering the right of rectification not possible in practice, should then the entire AI model, including the personal data in question, be deleted? As this example shows, it is still difficult to fully appreciate both the privacy risks and the consequences of the application of privacy laws to AI models in the current state of the art.

Box 1.1. Real and potential risks associated with AI systems

The OECD has worked to identify real and potential risks associated with AI systems, including generative AI, across its workstreams. Some risks are listed below:

- The amplification of mis- and dis-information at a large scale and scope, particularly through creation of artificial content that humans mistake for real content;
- AI model “hallucinations” that give incorrect or non-factual responses in a credible way, or the generation of illicit images such as fake child sexual exploitation material (e.g. “fake nudes”) ;
- Harmful bias and discrimination at an increased scale;
- Risks to privacy and data governance, at the level of training data, at the model level, at the intersection of data and model levels, or at the human-AI interaction level;
- Challenges to transparency and explainability due to the opacity and complexity of large models;
- The inability to challenge the outcome of models; and,
- Privacy breaches through the leaking or inferring of private information, among others.

Sources: (OECD, 2023[13]); (OECD, 2023[14]); (Lorenz, Perset and Berryhill, 2023[12]).

Some risks from generative AI are poorly understood but would be very harmful if they materialised, for example, leading to systemic, delayed harms such as embedded and perpetuated bias and labour disruptions among others, and to collective disempowerment (Lorenz, Perset and Berryhill, 2023[12]). Examples include models that display negative sentiment towards social groups, link occupations to gender (Weidinger, 2022[33]) or express bias regarding specific religions (Abid, 2021[34]). While some risks are identified as explicitly relating to privacy, some risks touch on topics that the privacy community already addresses without these risks necessarily being identified as involving "privacy” by the AI community (e.g. explainability, transparency, self-determination, challenging the output of an automated decision-making process, etc.). Before delving into these issues, it is important to acknowledge that while measures and instruments from both the AI and privacy communities can help mitigate known harms caused by the use of AI systems, they may have limitations in addressing intentional malicious uses of the technology. This highlights the necessity for broader collaboration between both communities to investigate, prevent, and mitigate potential misuses of AI.

Privacy concerns emerging from generative AI: Privacy Enforcement Authorities step in

The privacy and data protection issues raised by generative AI have quickly become a core area of focus for many PEAs. PEAs address these concerns either at the national level or in the context of the regional or global co-operation networks which are now part of their daily operations. Three recent initiatives of networks of privacy regulators must be specifically noted here, whereas an overview of national and regional initiatives is provided further down (section IV).

G7 Roundtable of Data Protection and Privacy Authorities

In June 2023 the G7 Roundtable of Data Protection and Privacy Authorities ("G7 DPA Roundtable") issued a statement on generative AI (G7, 2023[35]) listing key areas of concerns from a privacy and data protection perspective, which include:

- Legal authority for the processing of personal information, particularly that of minors and children, in relation to train models;
- Security safeguards to protect against threats and attacks that can leak personal information originally processed in the database used to train the model;
- Mitigation and monitoring measures to ensure personal information generated by generative AI tools is accurate and non-discriminatory;
- Transparency measures to promote openness and explainability in the operation of generative AI tools;
- Technical and organisational measures to ensure the ability for individuals affected by or interacting with these systems to exercise their rights, e.g. to erasure or not to be subject to automated decisions;
- Accountability measures to ensure appropriate levels of responsibility among actors in the AI supply chain;
- Limiting collection of personal data to only that which is necessary to fulfil the specified task.

The G7 DPAs urged close attention by technology companies to legal requirements and guidance from the DPAs when developing AI systems and services that use generative AI in Italy. The G7 DPAs also stressed that privacy and other human rights must be recognised and protected by those who design, develop, and deploy AI products and services, including generative AI. Shortly after, on 29 January 2024, following its fact-finding efforts, the Italian PEA formally notified the company of the existence of GDPR breaches (Garante per la protezione dei dati personali, 2024[36]). The Office of the Privacy Commissioner of Canada (OPC), the Office of the Information and Privacy Commissioner for British Columbia, the Commission d'accès à l'information du Québec, and the Office of the Information and Privacy Commissioner of Alberta are also jointly conducting an investigation into OpenAI's ChatGPT (Office of the Privacy Commissioner of Canada, 2023[37]).

Web scraping statement of the GPA's International Enforcement Co-operation Working Group

On 24 August 2023, twelve international PEAs, members of the International Enforcement Co-operation Working Group (IEWG) of the Global Privacy Assembly (GPA), adopted a joint statement to address the issue of data scraping on social media platforms and other publicly accessible sites (GPA's International Enforcement Cooperation Working Group, 2023[28]). The group notes that web scraping raises significant privacy concerns as these technologies can be exploited for purposes including monetisation through reselling data to third-party websites, including to malicious actors, private analysis or intelligence gathering. The joint statement: i) outlines the key privacy risks associated with data scraping; ii) sets out how social media companies (SMCs) and other websites should protect individuals' personal information from unlawful data scraping to meet regulatory expectations; and iii) sets out steps that individuals can take to minimise the privacy risks from data scraping.

Global Privacy Assembly's Resolution on Generative Artificial Intelligence Systems

These concerns were reiterated by the 45th Session of the GPA on 20 October 2023 in a Resolution on Generative Artificial Intelligence Systems (GPA, 2023[15]).

In this landmark Resolution, the GPA endorses a series of data protection and privacy principles as core elements for the development, operation, and deployment of generative AI systems, including:

- Lawful basis for processing;
- Purpose specification and use limitation;
- Data minimisation;
- Accuracy;
- Transparency;
- Security;
- Privacy by Design and Default;
- Rights of data subjects;
- Accountability.

Among other findings, the GPA emphasised the difficulty of reconciling the data minimisation and purpose limitation principles, which are core elements of privacy policy globally, with massive and indiscriminate collection of training data for machine learning.

Generative AI enhances the urgency to work on the interplay between AI and privacy regulations

As seen above, the community of PEAs has been active when it comes to responding to privacy risks in generative AI, emphasising an urgent need to reflect on the interplay between privacy frameworks and rules on trustworthy AI in general.

Against the backdrop of actions from both the privacy and data protection community on the one hand, and the rise in proposed regulations of AI like the EU AI Act on the other, comes the question of how these frameworks will fit together, namely as proposed AI legislation will enter amid well-established privacy and data protection rules and enforcement action by regulators. Co-regulatory approaches could be envisioned, when appropriate, to avoid the duplications and overlaps of requirements under different regulatory frameworks. In the EU specifically, the EU AI Act amplifies the need for clarity on the interplay of regulations with the General Data Protection Regulation (GDPR). For instance, GDPR's protections for individuals against forms of automated decision making (ADM) and profiling have been applied by courts and regulators alike for several years, ranging from detailed transparency obligations to applying the fairness principle to avoid situations of discrimination and strict conditions for valid consent in ADM cases (Barros Vale, 2022[38]) (Barros Vale, 2022[38]). Other studies illustrate how some PEAs take fundamental rights into account in their decisions, which is a relevant aspect of the assessment obligations outlined in many of the AI regulations currently under discussion (Grazia, 2017[39])

While such precedents present opportunities to enhance or clarify obligations under the AI Act through the lens of ADM jurisprudence, the overlap can also be a source of confusion for policy- and lawmakers and generates some unclarity regarding compliance. This is especially true for actors such as SMEs that are equipped with limited resources to understand the interplay of AI and data protection and privacy rules in practice. The numerous enforcement actions already undertaken by PEAs in the generative AI domain globally show that large portions of AI practices already fall under intense regulatory scrutiny, making it urgent to bring the two communities together, so as to ensure that the policy objectives underlying the different legislations are all satisfied.

At the same time, to fully unlock the potential of generative AI, substantial, diverse, and relevant data is essential for training models effectively. Having greater access to data enables AI models to perform better as they have the ability to learn from examples in an iterative process. In parallel, having varied and high-

quality data (e.g. accuracy, completeness, consistency, reliability, validity, timeliness) is key to building trustworthy algorithms and enhancing AI model performance. In this cycle, smooth and efficient data flows are crucial for optimal AI model functioning, ensuring a continuous exchange of information for ongoing learning and refinement of these models. In addition, the availability of training data from various sources, particularly from as many regions or countries as possible, is essential to contain the risk of the level of bias in AI systems, especially for models used across borders. In this regard, an increasingly frequently expressed concern is that more barriers to cross-border data flows are being adopted globally, covering both personal and non-personal data, raising a risk that data may become less accessible (or potentially restricted to specific regions or countries) for the development of AI-driven tools. These obstacles are not fundamentally new and have been analysed in part in the context of the work of the OECD to advance international policy discussions to harness the full potential of cross-border data flows under the banner of Data Free Flow with Trust (DFFT) (OECD, 2023[40]). For this, prioritizing international collaboration and aligning policy responses to foster trust and facilitate data exchange becomes essential. By capitalizing on commonalities and areas of convergence, the privacy and the AI communities can address obstacles that might impede the widespread, lawful, and successful deployment of AI.

Relatedly, for data to be available and have the most effective impact, it needs to be appropriately accessible. That means encouraging better coordination, as well as access to and sharing of data between organisations in the public sector and the private sector. These aspects have been addressed in the context of the work leading to the adoption of the OECD Recommendation on Enhanced Access and Sharing of Data (OECD, 2021[7]). In this context, an increasing number of initiatives are emerging to ensure data availability while protecting privacy, which are worth exploring. These include, for example, regulatory sandboxes and/or the involvement of data intermediaries (such as data trusts or data cooperatives) to foster responsible data sharing within the AI economy, which are provided in a growing number of privacy legislations (such as Singapore's PEA) and are the subject of complementary workstreams at the OECD.

2 Mapping existing OECD principles on privacy and on AI: key policy considerations

The OECD AI Principles set in the 2019 Recommendation on AI, and updated in 2024, can be used as an analytical grid for comparison with the privacy principles set in the landmark 1980 Privacy Guidelines, revised in 2013. The OECD AI principles are divided into two categories: 1) five value-based principles that serve as guidance for governments to develop AI strategies and policies for trustworthy AI, which can also be used by companies and AI developers; and 2) five recommendations to governments for national policies, for AI ecosystems to benefit societies (OECD, 2019[11]).

This mapping aims to identify priority areas of co-operation between the AI and privacy and data protection communities and begins by analysing key terminology used in the five values-based principles in the OECD Recommendation on AI. Analysis around the five recommendations to governments, which cover actions to foster vibrant AI ecosystems, could be covered in subsequent analysis. This mapping exercise permits the identification of:

- Policy areas where joint work between the two communities, at the OECD and beyond, could yield strong mutual benefits;
- Policy areas where synergies are low or non-existent; and
- Terminological differences between the two policy communities that could hinder interoperability and coordination.

This mapping exercise refers to established principles of privacy and data protection which have grown out of the OECD Privacy Guidelines. Some of these established principles, such as data minimisation, and rights with regard to automated decision making, do not appear explicitly termed as such in the OECD Privacy Guidelines, but have *de facto* become state-of-the-art in privacy policy through the implementation of the OECD Privacy Guidelines by OECD members.

Table 1. The OECD AI Principles, revised 2024

5 value-based principles for trustworthy, human-centric AI	5 recommendations to governments for AI ecosystems to benefit societies
1.1 Inclusive growth, sustainable development and well-being	2.1 Investing in AI research and development
1.2 Respect for the rule of law, human rights and democratic values, including fairness and privacy	2.2 Fostering an inclusive AI-enabling ecosystem (data, compute, technologies)
1.3 Transparency and explainability	2.3 Shaping an enabling interoperable governance and policy environment for AI
1.4 Robustness, security and safety	2.4 Building human capacity and preparing for labour market transformation
1.5 Accountability	2.5 International co-operation and measurement on trustworthy AI

Note: The 2019 OECD AI Principles were updated by OECD Ministerial Council in May 2024.

The methodological choice to map privacy considerations on to the OECD AI Principles in no way subsumes one framework to the other. Since this analysis focuses on privacy in the context of AI systems, the OECD AI Principles appeared as an appropriate starting point for comparative analysis. Moreover, an increasing number of data protection laws now include additional provisions relevant to AI (e.g. limitations on automated decision making, explicit incorporation of privacy by design/default principles) and must be considered when analysing the intersection of different frameworks at the national level.

The five values-based principles in the OECD AI Recommendation

The OECD AI Recommendation promotes the use of AI that is innovative and trustworthy and that respects human rights and democratic values, including privacy and data protection. The Recommendation on AI includes a definition of an AI system, which is currently used in AI frameworks around the world, such as the EU AI Act and the Council of Europe's Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law. According to this definition, an AI system is "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment" (OECD, Updated 2023[41]).

Key policy considerations from mapping AI and privacy principles

The OECD's work on classification and accountability of AI systems (OECD, 2022[17]; OECD, 2023[14]) has yielded a clear understanding of the AI system lifecycle, a framework that can help both privacy and AI communities analyse risks and mitigation measures in a uniform manner. On the privacy side, the privacy community has advanced methods for evaluating the impacts of AI on privacy rights in specific cases, and how different rights and collective interests, including the utility of AI systems for society, can be balanced in a manner respectful of privacy principles and the rule of law.

For certain generative AI risks, particularly those relating to systemic, delayed harms to society, or the risks of autonomous generative AI agents (Lorenz, Perset and Berryhill, 2023[12]), privacy and AI communities share a level of uncertainty regarding how to address such issues, and thus have a high interest in collaborating as concrete situations, risks, and policy solutions emerge. Table 2 outlines preliminary findings on the likely benefits and relevance from coordination between AI and privacy policy communities, based on the five values-based AI principles in the OECD Recommendation on AI.

Table 2. Overview of similarities and relevant areas of coordination between AI and privacy policy communities

OECD AI Principle	Preliminary analysis of similarities and relevant areas of coordination between AI and privacy policy communities
Principle 1.1 Inclusive growth, sustainable development and well-being	• Weighing economic and social benefits of AI against risks to privacy rights. • Promoting data stewardship for trustworthy AI in pursuit of beneficial outcomes for people and the planet where data controllers act with the benefits of data subjects in mind. • Coordination on considerations for data required to address environmental harms, displacement of labor, and other economic impacts (positive & negative) of AI.
Principle 1.2 Respect for the rule of law, human rights and democratic values,	• Identifying, assessing and treating AI risks to privacy and data protection. • Learning from each community's terminology. • The principles, concepts, and rules that structure the policy discourse in the privacy realm are important to know for the AI policy community; they include concepts like personal data/information, lawfulness of processing, the purpose and use limitation principles, personal data minimisation, privacy by design/by default, data retention, personal data accuracy, and data subjects' rights (e.g. individual control).

including fairness and privacy	• Existing alignment between AI data preparation for processing, which involves data cleaning and deduplication, to ensure the accuracy of the AI model and the data quality and accuracy privacy principles. • Harmonising requirements and methodologies of Human Rights Impact Assessments and Privacy Risk Assessments, which are a pillar of accountability under both approaches – when applicable.
Principle 1.3 Transparency and explainability	• Coordination on transparency and explainability to data subjects – the persons impacted by AI processing – which is a long-standing focus of privacy and data protection communities, in co-operation with consumer protection communities. • Interdisciplinary work on the connection between existing legal requirements for the explainability of AI systems set out in frameworks like EU GDPR and the current state of the art in the field of explainable AI.
Principle 1.4 Robustness, security and safety	• Clarification that both AI risk-based approach regulation and global data protection rules should be considered together to ensure the effective integration of privacy principles and considerations throughout the AI system lifecycle. • Coordination on data security (e.g. confidentiality, integrity) and Privacy Enhancing Technologies (PETs)
Principle 1.5 Accountability	• Incorporating AI system lifecycle methodology in privacy management programmes (PMP). • AI experts can leverage the work done in the field of privacy accountability by the privacy community, including at the OECD (see implementation guidance on the OECD Privacy Guidelines). • Harmonising oversight of Human Rights Impact Assessments and Privacy Risk Assessments – when applicable.

Note: This table represents preliminary analysis and is not exhaustive.

Differences in terminology between the two policy communities can hinder mutual understanding and further coordination. Table 3 provides a list of key concepts that often have different meanings between AI and privacy policy communities, in order to promote awareness of these possible variations, thus improve mutual understanding between the communities and optimise co-operation actions.

Table 3. Key concepts with different meanings between AI and privacy policy communities

Terminology	Preliminary analysis of differences in meaning, focus or implementation of key concepts
Fairness	For **AI policy communities**, fairness often refers to outcomes from the application of AI (such as predictions, recommendations, or decisions) that are based on algorithms and datasets with consideration for bias, for example, through mitigating algorithmic or dataset bias for specific groups (e.g. those categorised by class, gender, race, or sexual orientation). For **privacy policy communities**, fairness mainly refers to reasonable and transparent practices, respectful of consumers' and citizens' interests. It covers the prohibition of deceptive or misleading practices at the time of data collection, the obligation to handle people's data only in ways they would reasonably expect, and to consider how the processing may negatively affect the individuals concerned. Discrimination is one form of unfairness, but it is not the only one.
Transparency and explainability	For **AI policy communities**, transparency, explainability and interpretability have different meanings but overall refer to the good practice of AI actors providing accessible information to users to foster a general understanding of AI systems, making stakeholders aware of their interactions with AI systems, enabling those affected by an AI system to understand the outcome, and to enable those affected by an AI system to challenge its outcome based on plain and easy-to-understand information that served as the basis for the prediction, recommendation or decision. For **privacy policy communities**, transparency is a positive legal obligation to inform individuals, from whom personal data are collected, no later than at the time of data collection, on the use purposes for which consent is requested and the subsequent use is then limited to the fulfilment of those purposes. It also includes notifying these individuals about the data collection and use purposes when the processing of the data is based on another legal basis, so that individuals may consequently exercise their privacy rights, permitting them to exercise human agency, including free choice, optionality and redress.
Privacy and data protection	While the **AI policy communities** recognise privacy as a human right, "privacy" and "data protection" are commonly used in a narrower sense to refer to personal information included in datasets for training AI, and to the risks related to the loss of personal data through leakage or inference by AI models/systems. For **privacy policy communities**, privacy and data protection are part of the larger, overall human rights and consumer protection legal fabric, covering threats to fairness, lack of transparency, and threats to data security and robustness, to be addressed through data subjects' rights, accountability, and regulatory intervention. From this perspective, privacy is not only an individual right but also a social value.

Note: This table represents preliminary analysis and is not exhaustive.

Overview of possible commonalities and divergences in AI and privacy principles

Principle 1.1: Inclusive growth, sustainable development and well-being

Sometimes referred to as "People and Planet", this category covers a wide range of interests and risks, including harms to the environment, impact on jobs, and harms to vulnerable populations. In addition to the risks, Principle 1.1 focuses on the positive effects of AI on society, including improving healthcare and fighting climate change.

While collaboration between privacy and AI communities can lead to positive environmental outcomes, such as smart cities initiatives like traffic optimisation or waste management, there appears to be limited synergies for collaboration between privacy and AI policy communities when it comes to protection of the environment. While crucial, environmental protection largely falls outside the scope of privacy and data protection regulation, except when cross-referencing demographic data with environmental data to understand the impact of climate change on various groups. A discussion on the overlaps between data protection, freedoms and the environment seems to be underway, for the moment in a more prospective than normative manner (CNIL, 2023[42]). As well, the Global Privacy Assembly for instance mentions environmental harms as among those caused by indiscriminate collection of training data in violation of the data minimisation principle (GPA, 2023[15]).

The positive effects of AI on economic well-being, for example by decreasing the costs of products and services through, for instance, the automation of specific tasks, or improving health-related outcomes, are also not the direct focus of data protection regulation. Nevertheless, achieving these benefits can require striking a balance between protection of privacy and other human rights. For example, increased accuracy of AI systems will create better predictions leading to better health outcomes. But higher accuracy may come with trade-offs with privacy data protection rights as greater amounts and higher quality training data are needed. PEAs and courts have been dealing with these trade-offs for many years, applying constitutional mechanisms to resolve conflicts between competing rights and interests in specific cases, using, for example, the "proportionality test" in some OECD member countries. This balancing exercise is also recognised in certain legislations. The GDPR (Art. 1(3)), for instance, notes that "the free movement of personal data should not be restricted nor prohibited for reasons connected with the protection of natural persons with regards to the processing of personal data" (European Union, 2016[43]).

Synergies appear high between privacy and AI communities in discussing how different collective interests, such as better public health or security resulting from AI, can be balanced against increased interference stemming from AI systems with certain human rights. A prime example of this pathway is the collective effort to leverage data-driven tools, including AI, in the fight against the COVID-19 pandemic while respecting data protection and privacy principles.

Principle 1.1 also includes harms to vulnerable populations. Economic displacement is not a direct focus of data protection law and the effect of AI on jobs is outside the direct scope of data protection regulation. However, common privacy tools such as privacy management programmes (PMP) require data controllers to develop appropriate safeguards based on privacy risk assessment, and "risk" is intended to be a broad concept, taking into account a wide range of possible harms to individuals (OECD, 2023[44]). As well, the protection of vulnerable groups' personal information, in particular children, is a core focus of the WPDGP (OECD, 2021[45]) (OECD, 2021[46]) (OECD, 2021[47]) (OECD, 2022[48]). It is furthermore anticipated that the concept of digital vulnerability will be further discussed with a prospective outlook. This is the case of the monitoring of the elderly, the disabled, or patients, for instance as brain computer interfaces (BCIs) are being used to modulate brain activity for cognitive disorder management, or in relation to the increased use of smart wearable devices to monitor and detect occupational physical fatigue of employees in the workplace.

Key policy consideration: Principle 1.1 - Inclusive growth, sustainable development and well-being

Most of the interests and risks covered by Principle 1.1 are not the core focus of privacy and data protection laws. However, when applying privacy and data protection laws, courts and authorities often consider the social benefit of an AI application, including in improving public health or security, when evaluating the proportionality of its interference with privacy and data protection rights. Co-operation between the two communities could be focused on how to provide guidance on balancing AI social benefits and risks, including risks to privacy rights.

Principle 1.2: Respect for the rule of law, human rights and democratic values, including fairness and privacy

Principle 1.2 is divided into three sub-categories: bias and discrimination, privacy and data governance, and human rights and democratic values (OECD, 2023[14]).

Undue bias and discrimination

Bias and discrimination are important risks associated with AI, studied by both the AI community and the privacy community. Generative AI, as currently developed, deployed and used in the absence of guardrails, has amplified these risks due to the massive scale and scope of the application of such systems, and their input training data during their development phase. While the objectives are aligned between the AI community and the privacy policy community, the way bias and discrimination are studied in the two communities differs.

Bias and discrimination as studied by the AI community

Bias and discrimination are heavily studied in AI policy circles, and are cited as major concerns for generative AI, due to the increase in scale and scope of potential algorithmic bias resulting from foundation models and the massive training data they use (OECD, 2023[13]; Lorenz, Perset and Berryhill, 2023[12]). The work of the AI community has brought to light the many sources of algorithmic bias (OECD, 2023[14]), including: historical bias, representations bias, measurement bias, methodological and evaluation bias, monitoring bias and skewed samples, feedback loops and popularity bias (OECD, 2023[14]). AI scholars have also identified the characteristics and incompatibilities between different forms of fairness or non-discrimination, including equality of opportunity, equality of outcome or statistical parity, and counterfactual justice (OECD, 2023[14]): due to these incompatibilities, creating an “un-biased” AI system is extremely challenging.

Bias and discrimination as studied by the privacy community

PEAs examine illegal discrimination as one of the negative effects that may render the processing of personal data unfair, and thus unlawful. The lawfulness of processing personal data is evaluated in light of, among other things, the risk of discrimination that may ensue. The concern for discrimination is raised as one of the most important issues for PEAs when addressing the enforcement challenges posed by emerging technologies (OECD, 2021, p. 48[9]) .

To prevent discrimination based on the use of personal data, in many countries certain types of personal data have been designated sensitive, and therefore their permitted uses may be more limited and even

prohibited. This may be the case with age discrimination in employment decisions, gender discrimination in credit decisions, or political affiliation in government services and allowance allocation.

Trustworthy AI requires trustworthy data. The quality of the datasets used is essential for the optimal performance of AI systems. Throughout the data collection process, there is a potential for the inclusion of socially constructed biases, inaccuracies, errors, over- or under-representation, and mistakes. Consequently, there is a very high incentive for both AI actors and privacy advocates to seek high-quality datasets: AI actors are keen to work with the most accurate data possible to maintain trustworthiness in the outcomes and high model performance, while individuals strive to ensure that the AI model will not produce negative outcomes based on incorrect, incomplete, or insufficient data. In the OECD Privacy Guidelines, the data quality principle refers to the relevance of personal data in relation to the purposes for which they will be used. The more relevant the data, the higher its quality. Data relevance is crucial for generative AI tools, particularly concerning the potential inclusion of exogenous false or misleading information in the data (OECD, 2023[49]). Accuracy, completeness and timeliness are also important elements of the data quality concept. Complete and up-to-date information not only increases the quality of the data but also improves its accuracy, mitigating the risk of harmful bias.

Privacy management programmes (PMPs) include risk assessments that evaluate, among other things, the probability and potential impact of discrimination (OECD, 2023[44]). In jurisdictions, such as the United States that address privacy and data protection through the lens of consumer protection, the discriminatory effects of an AI system could result in a finding that the system is an “unfair practice” prohibited by law (FTC, 2022[50]). Discrimination is also addressed by other, non-privacy, laws, including employment, education, and banking laws, which complement the legal data protection landscape. Nevertheless, most data protection enforcement authorities interpret their role as being, at least in part, to prevent discrimination, including outcomes resulting from AI systems that process personal data.

Because of the different but complementary focuses of the AI community and the privacy policy community around discrimination, coordination between both communities on the discriminatory effects of AI systems would be highly beneficial. Privacy policy communities can learn the many nuanced approaches to non-discrimination, such as fairness considerations, being studied by AI experts, while AI policy communities can benefit from the experience of data protection communities in evaluating and sanctioning data controllers whose systems generate discriminatory outcomes. The possibility for synergies are thus high in combining the AI community’s expertise in approaches to bias with the privacy communities’ experience in evaluating discriminatory effects of AI systems in concrete cases.

Terminology differences: different meanings of fairness

Optimal coordination of privacy and AI approaches to non-discrimination can be hampered by definitional and terminological issues.

Fairness in AI discourse

For AI communities, fairness is often understood to refer to outcomes from the application of AI (such as predictions, recommendations, or decisions) that are based on algorithms and datasets with consideration for bias, for example, through eliminating algorithmic or dataset bias for specific groups (e.g. those categorised by class, gender, race, or sexual orientation). Bias is a systematic (as opposed to a random) error, associated with certain categories of data inputs. For example, a facial recognition algorithm may make more errors for people wearing glasses than for people without glasses. A difference in error rates between people who wear glasses versus people without glasses can be problematic from a technical and operational perspective and might create discrimination or inequal treatment, but would not generally raise legal or ethical concerns because glasses are not a protected attribute. By contrast, a difference in error rates between men and women, or between people with light skin and dark skin, would be considered

unfair, because the bias (systematic error) is linked to an attribute which in the particular legal and cultural context is associated with population groups that have been historically disadvantaged. Protected attributes may be absent from the input data, but inferred from other apparently neutral attributes. For example, a postal code may become a proxy for ethnic origin if the postal code corresponds to a neighborhood where many inhabitants share the same ethnic origin.

The work on fairness in the AI community has yielded multiple identified sources of bias (systemic, computational and statistical, human-cognitive) (Tabassi, 2023[51]) as well as a realisation that bias can almost never be completely eliminated. Attempts by data scientists to transform fairness into more mathematical properties has helped open debates among legal scholars on non-discrimination laws, in particular on how discrimination should be measured and what should be considered illegal discrimination (Wachter, 2022[52]). Thus, a link exists between AI "fairness" and legal debates on non-discrimination laws.

Further confusion arises when fairness in the AI context is translated into different languages. In AI communities, the French term for fairness is "équité", which also means "equity" in English. Yet equity in English does not just mean absence of discrimination. It also means conduct that respects good faith and the spirit, not just the letter, of the law. Equity in English can also refer to equal opportunity for disadvantaged groups, allowing in some cases for compensating measures to redress structural disadvantages. Equity, which allows for compensating measures, is often contrasted with equality, which refers to strict equal treatment for each individual and hence without compensating measures. From an AI standpoint, both equity and equality are variants of fairness in the non-discrimination sense.

Fairness in privacy discourse

In the privacy policy community, fairness is conduct that is consistent with reasonable expectations. Often fairness is best defined by what it is not, i.e. "unfair" practices. Unfair practices are generally illegal under consumer protection law, competition law, privacy law, and non-discrimination law. Unfair practices may involve deceit and lack of transparency (Malgieri, 2020[53]). For the privacy policy community, transparency is a necessary element of fairness.

The prohibition of unfair practices may also target imbalances in economic power. This is because fairness in privacy is not solely concerned with the mathematical distribution of resources or outcomes but also considers context (including human decision-making) and other qualitative aspects, such as power imbalances between individuals and those who process their data (ICO, 2023[54]).Therefore, fairness may require imposing extra duties on more powerful actors to ensure that their dealings with less powerful actors reflect more balance (Clifford and Ausloos, 2018[55]). Linked to power imbalances is the individual participation principle of the OECD Privacy Guidelines, which empowers individuals to challenge data relating to them. In certain countries, "procedural fairness", sometimes referred to as "due process", may also impose procedural constraints on government actions. (Mulligan et al., 2019[56]) This could entail the right to discuss, or challenge, a particular aspect of data processing with a human representative of the other party, and eventually appeal the matter to an independent and impartial decision-maker. (Mulligan et al., 2019[56]) In this respect, fairness is linked to respect for human rights such as human dignity, individual autonomy, optionality, and redress.

In matters of privacy and data protection, fairness is sometimes equated to processing that is consistent with the reasonable expectations of the data subject, i.e. processing that would not "surprise" the data subject (Malgieri, 2020[53]). It requires data controllers to handle personal data in a manner that aligns with individuals' reasonable expectations and avoid using it in any way that could adversely affect them (Datatilsynet, 2018[57]). Practices that would lead to unlawful discrimination would also be considered unfair (CNIL, 2017[58]). Fairness is also associated with the concept of "good faith" (Malgieri, 2020[53]) and "loyauté" – loyalty – in French. Loyalty and good faith require avoiding conduct that violates ethical norms, and respecting the spirit, not just the letter, of laws (Mulligan et al., 2019[56]; Malgieri, 2020[53]).

To promote mutual understanding, AI and privacy communities should be aware that definitions of "fairness" vary between them.

Key terminology and concepts: privacy and data governance

There is the need for, and benefits of, coordination between AI and privacy communities on the concepts of "privacy and data governance" in the AI Principles. This element of human-centred values and fairness targets core data protection and privacy issues (OECD, 2023[14]). Previous OECD analysis highlights the risk that "AI systems can cause or exacerbate impacts of asymmetries in power and access to information, such as between employers and employees, businesses and consumers, or governments and citizens" (OECD, 2023[14]). Such access to information and related impacts of asymmetries has links to respect for privacy rights.

Some of the most significant generative AI risks come from systemic impacts on society, such as the generation of inaccurate synthetic content that may influence individuals' preferences, attitudes, and behaviors (Lorenz, Perset and Berryhill, 2023[12]). While the privacy community has traditionally focused on harms to individuals, increasing attention in the privacy community is turning to the "collective societal impacts of the use of mass amounts of personal data processed by emerging technologies" (OECD, 2021[9]).

Coordination between AI and privacy policy communities would help clarify the role of data protection and privacy laws and of Privacy Enforcement Authorities in addressing both individual and collective harms created by generative AI. Today, privacy and data protection practices reflect principles in the OECD Privacy Guidelines, but also principles that have emerged from the Guidelines as best practices with respect to privacy considerations among OECD members. Those principles include, in particular: lawful basis for processing; purpose specification and use limitation; data minimisation; accuracy; transparency; security; privacy by design and default; rights of data subjects, including with regard to automated decisions; and accountability.

Building trustworthy AI requires trust in different aspects of data use, such as acquiring reliable data, processing it lawfully, using it responsibly, keeping it safe, and being transparent about their uses. A key step is to ensure that AI systems are lawful at every stage of their lifecycle. In this regard, most privacy and personal data protection frameworks require that there be a "lawful basis" for both collecting and processing data. While most laws generally provide for a series of such legal bases, in practice, the legal basis known as "legitimate interests" is the one which is considered the most suitable in the context of Generative AI. This requires that the interest pursued by the AI developer, provider or user (e.g. in developing or implementing a model) be legitimate, that the data processing at stake is effectively needed to meet this legitimate interest, and that it does not create disproportionate interference with the interests and rights of data subjects. As mentioned earlier, striking the right balance between these different interests can be complex in the current state of the art and calls for reinforced co-operation between the AI and the privacy community.

Privacy laws also provide that only the minimum amount of personal data necessary should be processed to achieve the intended purpose. This so-called data minimisation principle is implicit in the OECD Privacy Guidelines and made explicit in various privacy laws, such as the GDPR or the California Privacy Rights Act. AI business models, especially in the wake of generative AI, have followed the assumption that collecting extensive amounts of data is essential for the effective operation of AI systems, especially during the training phase. This approach may conflict with the data minimisation principle, since it may not be possible to map in advance what personal data the AI system requires.

Yet the concept of data minimisation does not mean either to completely avoid processing personal data or limiting the amount of data to a specific volume (ICO, 2023[59]). The application of this principle should be contextualised, considering the concrete AI system and its objectives. For proper contextualisation, it

is essential to identify the purposes for which the data will be used before, and in any case, not later than at the time of data collection (OECD, 2023[60]). This involves not only respecting regulatory requirements, such as specifying the purposes to the data subjects, but also considering ways in which the same purposes could be achieved with less data. Sometimes, in the context of AI, data minimisation may be more effectively achieved by prioritising data quality over quantity. For example, identifying whether the AI model can be trained without the use of sensitive personal data by utilising an existing public data source. This is because the quality of the training data can have a more significant impact on model accuracy than the quantity of the training data. There are also several techniques that can be adopted in order to develop AI systems that process only the necessary data while still achieving performance objectives (OECD, 2023[61]). There are also proposals to make the data used to train AI models openly accessible (not just the model's code or architecture), allowing for scientific evaluation of the data feeding the models to facilitate the development of less data-intensive models (Widder, 2023[62]).

PEAs play a crucial role in interpreting and guiding the application of privacy considerations in the context of AI. This role is particularly relevant when integrating key privacy concepts, such as the purpose specification principle, with the unique characteristics of AI. An example of this can be seen in the CNIL's practical guidelines ("AI how-to sheets"), which provide insights to organisations seeking to define the purpose(s) of AI systems in alignment with EU GDPR, while also considering the specific nuances of their development process. In this regard, the CNIL differentiates between two scenarios: one where the operational use of the AI system is clearly identified during the development phase, and another where it emerges during the deployment phase. In cases where the operational use of AI systems is clearly defined during development and remains consistent in deployment, processing operations in both stages typically serve a single overarching purpose. Therefore, if the purpose in the deployment phase is specified, explicit and legitimate, it is understood that the purpose in the development phase shares these qualities (CNIL, 2023[63]). The operational use of AI systems during deployment is not always clearly outlined during development, however, especially in the case of general-purpose AI systems such as foundation models. In this scenario, the purpose of developmental processing is deemed determined, explicit, and legitimate only if it is adequately precise. The CNIL's guidelines offer examples to illustrate when a purpose is considered sufficiently precise. (CNIL, 2023[63])

Key terminology: human rights and democratic values

Human rights include civil and political rights such as equality, non-discrimination, freedom of expression and association, privacy, and economic, social, and cultural rights such as education or health (OECD, 2023, p. 31[14]). While AI can mitigate some barriers to equal access to opportunities and services, such as AI-based translation that can improve access to healthcare or AI assistive technologies for people with disabilities to perform daily tasks or integrate the labour market, AI can create risks at a macro-level, such as disrupting balances of power through manipulation and polarisation of opinions at scale (OECD, 2023[14]). Generative AI increases these risks, particularly for massive manipulation of public opinion for example through misleading synthetic content (OECD, 2023[13]; OECD, 2023[8]). Other human rights impacts include:

- Impact on human agency and self-determination (OECD, 2022[17]);
- Impacts on freedom of thought, freedom of expression, non-discrimination, the presumption of innocence and the right to a fair trial (OECD, 2022[17]);
- Impacts on the ability to access key services, such as education, healthcare or banking services (OECD, 2022[17]);
- Impacts and negative externalities for vulnerable populations (in particular children and disadvantaged groups) (OECD, 2023[14]); and,
- Asymmetries in power and access to information (OECD, 2023[14]).

Guarantees to these human rights are typically covered by numerous, non-privacy, laws. However, as pointed out in the preceding section, privacy and data protection laws, and privacy risk assessments, are part of the overall human rights fabric, as an infringement of the right to privacy may cause and lead to infringements of additional human rights, and these risks must be taken into account (OECD, 2023[44]). The AI community increasingly relies on Human Rights Impact Assessments (HRIA) (OECD, 2023[14]), whereas the privacy community relies on "Privacy Risk Assessments" (OECD, 2023[44]), "Privacy Impact Assessments" (GPA, 2023[15]) or "Data Protection Impact Assessments" (GDPR).

The Chapter on Accountability of the OECD Privacy Guidelines Implementation Guidance refers specifically to risks related to, for example, an individual's eligibility for a right, privilege, or benefit; psychological manipulation; and impact on vulnerable populations (OECD, 2023[44]). Under existing privacy and data protection laws, courts and PEAs are already evaluating AI systems' impacts on privacy rights to determine whether an AI system is fair, lawful, and proportionate. More specifically, courts or PEAs weigh the interferences with privacy rights against the social benefits derived from using an AI system. Recent case law of the Court of Justice of the European Union (CJEU) on data protection has yielded detailed analysis on how the right to public security should be balanced against privacy rights in the context of algorithmic systems designed to detect terrorist threats (namely, the CJEU Quadrature du Net case (Joined Cases C-511/18, C-512/18 and C-520/18) and La Ligue des Droits Humains case (Case C-817/19)). In these cases, the CJEU provided concrete guidance on the levels of interference with privacy and freedom of expression that are permitted in the context of protecting against risks of terrorism, as well as opining on specific risk mitigation measures, such as the level of human control that must accompany any deployment of such a system. The Court's analysis was based the GDPR and the European Union's Charter proportionality test.

Key policy consideration: Principle 1.2 - Respect for the rule of law, human rights and democratic values, including fairness and privacy

All aspects of Principle 1.2 merit close coordination between privacy and AI policy and oversight and enforcement authorities. In particular:

- Fairness discussions should be shared between AI and privacy communities because of differences in meanings of fairness that need to be understood by both communities. Fairness aims to eliminate unjustified adverse outcomes for individuals. This can be enhanced by considering the socio-legal perspectives of fairness in the privacy community, as well as the computational and governance perspectives from the AI community. By harmonising these perspectives, the implementation of the concept of fairness can be strengthened and tailored to address the unique risks that arise from AI. In this regard, the Catalogue of Tools and Metrics for Trustworthy AI on the OECD.AI Observatory (OECD.AI, 2024[64]) and the experience of privacy authorities in evaluating discrimination risks of AI systems in concrete cases can enrich such policy discussions on fairness in both communities.
- Privacy and data protection practices can further be clarified for the benefit of AI communities, for example in terms of their broad scope and implementation practices. This involves looking at the implications of the principles of data minimisation, the purpose principle, information and rights of individuals, among others. Joint efforts are needed to craft comprehensive governance and practical solutions for compliance, especially for small and medium-sized enterprises that can face challenges with complying with privacy obligations.

- Human rights and democratic values merit coordination because privacy law has extensive experience in managing trade-offs in competing rights and interests, which can be useful in concrete AI use cases. Human rights impact assessments and privacy risk assessments in the AI and privacy domains, respectively, could be brought closer together to avoid different, potentially conflicting, approaches.

Principle 1.3: Transparency and explainability

Transparency and traceability

Transparency describes responsible disclosure to ensure people are aware that AI is being used in a prediction, recommendation or decision, or in an interaction (e.g. a chatbot) (OECD, 2023, p. 33[14]). Important disclosures, such as making persons aware that they are dealing with AI or that their personal data are being used, is a common objective of both AI and privacy regulation, and increasingly discussed in the context of generative AI where chatbots have emerged with very popular applications across sectors. Where is close alignment in the description of transparency provided in (OECD, 2023[14]) in connection with AI accountability and relevant privacy laws (e.g. see [C(2021)42] para. 26; (GPA, 2023[15]), point 5).

Privacy authorities, as well as consumer protection authorities, have considerable experience in evaluating the sufficiency of information communicated to the data subject. Data protection laws like the GDPR mandate that information provided to individuals must be in a transparent, concise, intelligible, and easily accessible form, using clear and plain language, in alignment with the Openness Principle in the OECD Privacy Guidelines [OECD/LEGAL/0188]. Building on this approach, the recent Advisory guidelines on use of personal data in AI recommender and decision systems issued by the Singapore Data Protection Commission encourage the provision of information to users on several key aspects: (a) the function of the product that requires collection and processing of personal data (e.g. recommendation of movies); (b) a general description of types of personal data that will be collected and processed (e.g. movie viewing history); (c) explanation of how the processing of personal data collected is relevant to the product feature (e.g. analysis of users' viewing history to make movie recommendations); and (d) identification of specific features of personal data that are more likely to influence the product feature (e.g. whether movie was viewed completely, viewed multiple times) (Singapore Data Protection Commission, 2024[65]).

It can be a challenge to make highly complex data processing information understandable to individuals, particularly since many simply click through the transparency notification information as quickly as possible. This challenge is further complicated by the "black box" problem – where AI developers are unable to fully explain how an AI system came to generate an output – and the trade-off between accuracy and interpretability of AI models. The overarching objective of transparency requirements, however, is to guarantee individual participation, and human agency to consent to disclosure or control of facets of their data and related identities (e.g. body, data, reputation). This objective aligns with the objective of transparency for AI (Tabassi, 2023[51]).

In the data protection space one solution was introduced by application stores which require application developers who wish to be included in the store to provide concise details in the download screen on the developer's privacy policy, such as 'Data Used to Track You', 'Data Linked to You' and 'Data Not Linked to You' as well as a link to the developer's privacy policy. Initiatives and frameworks that have been suggested for enhancing transparency in the AI context can also follow similar methods. One such example is the use of "model cards" to report essential information about the characteristics of machine learning models. These cards can encompass a comprehensive range of metrics, evaluating bias, fairness, and inclusivity aspects (Margaret Mitchell, 2019[66]), alongside providing insights into the provenance of the

data, details of the statistical distribution of various factors in the training data sets, and other details on the data sets used in the creation of the model data's origin, statistical distribution of pertinent factors within the training datasets, and additional details pertinent to the datasets utilised in constructing the model. To enhance transparency in AI systems, organisations can establish dedicated organisational roles and functions, develop new policies and procedures, or revise existing ones. Additionally, documenting each stage of the design and deployment process of AI systems can facilitate the provision of meaningful explanations to affected individuals (ICO, 2020[67]).

Traceability in AI describes the need to maintain a complete account of the provenance of data, processes, code, and other elements in the development of an AI system (OECD, 2023, p. 33[14]). Highlighted in the OECD AI Principles, traceability has also been a long-time pillar of privacy accountability and management programmes. As pointed out in the Chapter on Accountability of the OECD Privacy Guidelines Implementation Guidance (OECD, 2023[44]) and in the *OECD Advancing Accountability in AI* report, "accountability comprises the taking of *responsibility* for personal data use and a means to *demonstrate* (or answer, verify, or make visible) this to other stakeholders", including the keeping of records of data processing (OECD, 2023, p. 33[14]).

AI traceability involves documentation regarding the planning and design of the system, and its testing (OECD, 2023, p. 33[14]). This goes beyond the scope of traceability for purposes of privacy and data protection, which is limited to documenting the steps in processing personal data. Nevertheless, the two overlap in significant ways.

AI systems, including generative AI, can challenge traceability because underlying foundation models are trained on massive amounts of data, which may include personal data that is unknown to the downstream user of the system, for example if a foundation model is subsequently fine-tuned Autonomous generative AI agents that continue to collect information when deployed in the field represent a particularly complex challenge to achieving traceability (OECD, 2023[13]). The unrestricted scraping of publicly available data for training of generative AI raises challenges for respect for other protected rights, such as intellectual property rights (OECD, 2023[13]), as well as for privacy and data protection laws.

Explainability and interpretability

Explainability and interpretability are present in policy discussions in both the AI community and the privacy community. The complexity and often "black box" nature of machine learning derived AI systems, and particularly generative AI models, make these issues increasingly urgent. Not only can black box challenges prevent users from understanding why a large model makes false statements or "hallucinations" (OECD, 2023[13]), such opacity can limit users' and designers' ability to understand and control unpredictable model behavior, leading to significant risks, including what some call "existential risks" which is still a subject of discussion in the scientific community (Lorenz, Perset and Berryhill, 2023[12]). Finding ways to pierce the veil of black box machine learning models or render those models more inherently explainable without sacrificing model performance, is an important and fast-moving field of research for data scientists. In a separate research stream, experts in human-computer interaction (HCI) focus on the effectiveness of explanations on people, taking into account their cognitive limitations.

The US National Institute of Standards and Technology (NIST) established four principles for explainable AI that constitute foundational characteristics for explainable AI systems. NIST proposes that explainable AI systems (i) deliver accompanying evidence or reasons for outcomes and processes ("explanation"); (ii) provide explanations that are understandable to individual users ("meaningful"); (iii) provide explanations that correctly reflect the system's process for generating the output ("explanation accuracy"); and (iv) that a system only operates under conditions for which it was designed and when it reaches sufficient confidence in its output ("knowledge limits") (NIST, 2021[68]).

The privacy community is increasingly focused on AI explainability to ensure accuracy, fairness, and accountability in data processing. Privacy and data protection laws and PEAs use the broad term "transparency", a term that for them subsumes explainability and interpretability of data processing practices. Courts and PEAs may impose explainability for algorithmic decisions that have important impacts on individuals, since without explainability, human oversight cannot be assured, and individuals cannot have an effective right of redress (CJEU, 2022[69]). Modern privacy laws have recognised transparency as the most effective technique for oversight and accountability. Arguably, some say that the real paradox of privacy could be that "privacy requires transparency" (Rotenberg, 2021, p. 497[70]), i.e., the protection of individual privacy might depend on systems and processes being open and transparent.

Many legal requirements exist for some form of explainability (Maxwell and Dumas, 2023[71]). Some of those requirements derive from privacy and data protection laws, others from fundamental rights texts or constitutional provisions guaranteeing due process. There is a clear link between explainability, human rights, and fairness. Some aspects of explainability and interpretability fall outside the natural remit of data protection law. For example, explainability and interpretability to help data scientists improve the model, or to help model developers and operators detect safety risks, would not be the main focus of data protection regulation. However, PEAs have extensively examined explainability and interpretability of data processing practices to help an individual affected by an algorithm understand the algorithm's decision and challenge it, or to help a system operator detect potential discrimination in the algorithm's output.

For example, in cases in which data subjects were subjected to automated decision-making or profiling in the sense of Article 22 GDPR, the Spanish Data Protection Agency (AEPD) has pointed out that data subjects should be able to understand the way in which the information concerning them will be processed (whether AI is involved, for example) and with meaningful information about the logic involved. The AEPD has highlighted that complying with this obligation by offering technical references on the implementation of the algorithm may be obscure, confusing or lead to information fatigue. Sufficient information should be provided in an accessible manner to enable the subjects to understand the behaviour of the processing. Although the type of AI component used will determine how these criteria should be applied, the AEPD's guide on AI-based data processing (AEPD, 2020[72]) includes examples of the types of information that may be relevant to the data subject:

- Details about the data used for decision-making beyond just the category, especially information regarding the duration of use of the data (how old the data are).
- Relative importance or weight given to each of the data in the decision-making.
- Quality of the training data and the type of models used.
- Profiling activities conducted and their implications.
- Error or precision values according to the specific metrics used to measure the validity of the inference.
- Existence or non-existence of qualified human supervision.
- References to audits, especially audits on possible deviations of inference results, as well as the certification or certifications of the [AI] system. For adaptive or evolutionary systems, the last audit conducted.
- If the [AI] system includes information referring to identifiable third parties, the prohibition of processing such information without legitimization and the consequences of doing so.

At the same time, it is worth noting that some PEAs (e.g. the UK ICO), have issued detailed guidance on explainability, and emphasised the importance of considering the audience for explanations beyond just the affected individual. This includes staff members who rely on the AI system to make decisions and need

to convey meaningful information to the affected individual, as well as auditors or external reviewers responsible for monitoring or overseeing the system's production and deployment (ICO, 2020[67]).

Terminological differences related to transparency and explainability

As noted above, AI communities may attach separate meanings to transparency, explainability and interpretability, whereas privacy communities will generally use the term transparency to include all those meanings, depending on the circumstances. Primary goals of explainability include ensuring that AI systems are understandable to those affected by their deployment, in addition to being understandable to their developers to help advance technological progress and improve model performance. Such objectives can be greatly enhanced by aligning the technical expertise within the AI community, incorporating techniques and tools that facilitate explainability, with the legal requirements for transparency and explainability stemming from privacy and data protection regulations.

Key policy consideration: Principle 1.3 - Transparency and explainability

Privacy and AI communities share common concerns for AI transparency, explainability and interpretability, even though the definitions and focuses of the two communities can differ.

Some areas of explainability and interpretability fall outside the scope of data protection legislation, for example when explainability is designed to help model designers improve the model performance. However, when it comes to helping persons affected by AI systems understand and contest their processes or outputs, or to help users detect algorithmic discrimination, data protection law and AI policy align.

The same alignment is evident in transparency and traceability. Transparency raises questions on how information on complex systems can be made easily accessible, understandable, and still complete, a subject long studied by privacy authorities. Traceability, is a common concern for AI and privacy policy communities where coordination should be encouraged.

Principle 1.4: Robustness, security and safety

Generative AI's complexity and opacity may complicate efforts to constrain model behaviour (Lorenz, Perset and Berryhill, 2023[12]), thereby hurting reliability. Homogenisation (Bommasani, 2021[73]) and "algorithmic monoculture" risk, which refers to the risk of running similar algorithms with similar vulnerabilities, (Kleinberg and Raghavan, 2021[74]) can make large models more vulnerable to a single defect or attack. A notable example of these challenges is a data breach experienced by ChatGPT in 2023, which exposed user information and prompt history (Kovacs, 2023[75]).

This principle broadly converges with the data security principle in the OECD Privacy Guidelines: "Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data" [OECD/LEGAL/0188]. Robustness, security and safety overlaps with privacy principles linked to accuracy and security highlighted in the Global Privacy Assembly's recent Resolution on Generative AI (GPA, 2023[15]).

Data protection aspects of security focus on harms to individual rights and freedoms, including consumer harm such as that arising from identity theft, as well as physical or emotional harm, for example when a data breach exposes information about a person's sexual identity or a woman's fertility status, which may

invoke emotional anxiety, social conflicts and physical harm in some instances and societies. Personal safety may also be at risk from government use of personal data such as political views or ethnicity. "Safety" is a growing area of analysis and concern in the AI policy community, and extends to aspects of physical safety resulting, for example, from a faulty medical diagnosis, a faulty autonomous vehicle, or a cyberattack on an AI-powered electric grid. Safety is generally the priority of cybersecurity regulation, combined with other, industry-specific, safety laws.

Robust machine learning models traditionally need large representative data sets for their training. This can conflict with the data minimisation principle (GPA, 2023[15]). Advances in machine learning that require less data, or that process data in protected ways – such as Privacy Enhancing Technologies (PETs) – can help reduce the gap between the development of safe AI models and the protection of individuals' privacy rights. However, more analysis is needed to understand how such considerations can be handled.

While traditional cyberattacks are enabled by vulnerabilities in software programs, their configuration and interface, attacks on AI systems are facilitated by inherent limitations in the underlying AI algorithms (Comiter, 2019[76]). Manipulating the input fed into the AI system to alter the output or corrupting the AI model itself, making it inherently flawed, are some examples (Comiter, 2019[76]). At the same time, empirical studies (Liwei Song, 2019[77]) show that privacy and security can sometimes be in tension: when efforts are made to increase machine learning models' robustness against adversarial attacks, it can sometimes render them more susceptible to risks such as the identification of individual's data. Similarly, recent research (Reza Shokri, 2019[78]) has demonstrated how some proposed methods to make machine learning models explainable can unintentionally make it easier to conduct privacy attacks on models. These problems may require supplementing existing frameworks with new approaches and solutions, including but not limited to those outlined in NIST's AI Risk Management Framework (NIST, 2023[79]) or the recent publication on Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST, 2024[80]).

Deception arising from generative AI models can also result in security vulnerabilities, especially when the AI system is employed in specific contexts, such as law enforcement, medicine, education, or employment. Governments have recently started to explore the use of new tools to identify and manage the security risks associated with generative AI. For example, the 2023 US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, emphasises the importance of "red teaming" – a process of identifying security risks through groups of individuals attempting to subvert security safeguards as a means to test AI systems - as a helpful testing process to identify vulnerabilities and flaws in AI systems (The White House, 2023[4]).

Key policy consideration: Principle 1.4 - Robustness, security and safety

There exists considerable alignment, and potential synergies, between AI and privacy policy communities on risks of data leakage from generative AI models, and on PETs. However, there is also a need for further understanding and co-operation between the two communities, particularly to support the AI community in an expanded understanding of the implications of long-standing privacy rules and standards in this area.

Principle 1.5: Accountability

Both privacy communities and AI communities have developed approaches to accountability and risk management. Although neither field invented accountability and risk management, which originate from the regulation of complex systems such as offshore drilling or financial services (Yeung, Howes and Pogrebna, 2020[81]), or from environmental protection law (e.g. the obligation to conduct impact assessments). Certain aspects of accountability come from compliance programs put in place by corporations to detect and prevent illegal activity by the corporation's employees and subcontractors, including corruption and competition law violations.

Both communities are investing considerable resources to ensure that risk management principles are effective for AI systems, and in particular those that use personal data. Within the OECD, both the WPDGP and WPAIGO have developed accountability frameworks (OECD, 2023[14]; OECD, 2023[44]), and there is further opportunity to connect both frameworks in the context of AI governance.

The work of the OECD on AI classification (OECD, 2022[17]) and accountability (OECD, 2023[14]) provides a detailed and robust framework for identifying risks, relevant stakeholders and mitigation measures for AI systems based on the AI system lifecycle. Work on responsible business conduct, for example through the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct.

From the privacy side, the OECD's work on accountability significantly elaborates on this concept found in the OECD Privacy Guidelines, and presents a framework for privacy risk management which is not AI-specific, in keeping with the approach of the Guidelines (OECD, 2023[44]). Further work bridging AI and privacy works could incorporate the AI lifecycle framework presented in OECD reports (OECD, 2022[17]; OECD, 2023[14]), and present in AI laws and regulations, for example in the EU AI Act.

In this context, there are already legislative efforts that seek to integrate privacy and data governance into the AI accountability framework. For instance, the EU AI Act sets out precise quality standards for datasets used in AI model training, such as the identification of relevant data gaps and biases that may impact human rights.

Key policy consideration: Principle 1.5 - Accountability

OECD frameworks for classification and accountability of AI systems rely on an AI system lifecycle model that can be integrated with privacy management programmes to facilitate consistency in methodological approaches.

3 National and regional developments on AI and privacy

There is a growing number of AI and privacy developments at the national and regional levels around the world. This section provides an overview of notable such initiatives at the intersection of AI, data protection, and privacy as early 2024. This list, enriched with contributions from OECD member and partner countries, does not claim to be exhaustive given the rapidity of developments in this area. This analysis highlights the diversity and complementarity of measures already adopted by various authorities., spanning positive incentives to more coercive actions.

International responses by Privacy Enforcement Authorities

PEAs are working together on responses to AI, including generative AI, including various statements and resolutions:

- Statement on Generative AI by the DPAs of G7 countries, adopted on 21 June 2023 (G7, 2023[35]);
- Resolution of the Global Privacy Assembly on Generative AI (Global Privacy Assembly, 2023[82]);
- Web scraping statement of 12 members of the GPA's International Enforcement Working Group (IEWG) (Global Privacy Assembly, 2023[83]).
- Resolution of the Global Privacy Assembly on Artificial Intelligence and Employment (Global Privacy Assembly, 2023[84]).

Guidance provided by Privacy Enforcement Authorities on the application of privacy laws to AI

PEAs have launched several initiatives and provided guidance (in certain cases with insights and clarifications gained from experiments conducted in regulatory sandboxes) in response to the growing use of AI technology, including, but not exclusively regarding, generative AI tools.

In December 2023, **Canadian** privacy regulators launched principles for responsible development and use of generative AI (Office of the Privacy Commissioner of Canada, 2023[85]). Federal, provincial and territorial privacy authorities developed together a set of principles that lay out how key privacy principles apply when developing, providing, or using generative AI models, tools, products and services.

In **France**, the CNIL created an AI department in January 2023 to strengthen its expertise on these systems and its understanding of the risks to privacy while anticipating the implementation of the EU AI Act. On 16 May 2023, the CNIL published its action plan for the deployment of AI systems that respect the privacy of individuals. The action plan builds upon the CNIL's previous efforts in the AI domain and comprises a series of activities aimed at supporting the deployment of AI systems that uphold individuals' privacy (CNIL, 2023[86]). The CNIL's action plan also includes a dedicated dossier on generative AI, highlighting the

technical functioning of generative AI, underlying legal issues and ethical challenges, and real-world applications. In April 2024, CNIL released “how-to” guidance for legal and technical professionals (data protection officers, legal professionals, people with AI-specific or non-specific technical skills, etc.) on the development of AI systems when it involves the processing of personal data (CNIL, 2024[87]).

The **Spanish** Data Protection Agency (AEPD) issued guidance on GDPR compliance regarding processes integrating AI (AEPD, 2020[72]) and on Audit Requirements for Personal Data Processing Activities involving AI (AEPD, 2021[88]). The AEPD has complemented this guidance with other initiatives, including a note that compares the differences between the concept of transparency in the EU AI Act and in the GDPR (AEPD, 2023[89]). The Ibero-American Data Protection Network released General Recommendations for the Processing of Personal Data in Artificial Intelligence (Ibero-American Data Protection Network, 2020[90]) and the Specific Guidelines for Compliance with the Principles and Rights that Govern the Protection of Personal Data in Artificial Intelligence Projects (Ibero-American Data Protection Network, 2020[91]).

The Personal Data Protection Authority (KVKK) of the **Republic of Türkiye** published "Guidelines on the Protection of Personal Data in the Field of Artificial Intelligence", which is based on a review of key international resources, including the OECD AI Principles. Tailored to the needs of developers, manufacturers, service providers and decision-makers in the field of AI, it provides a set of recommendations specifically designed to protect personal data in accordance with the principles set out in Law No. 6698 on the Protection of Personal Data. The main objective is to provide a comprehensive framework that ensures the responsible use of AI applications and promotes a safe and trustworthy environment for all stakeholders.

For example, the **United Kingdom’s** Information Commissioner’s Office (ICO) has produced comprehensive guidance on AI and data protection, that was updated on 15 March 2023 (ICO, 2023[54]). This piece is supplemented with specific guidance for explaining decisions made with AI (ICO, 2020[67]). Relatedly, on 15 January 2024, the ICO launched a consultation series on how aspects of data protection law should apply to the development and use of generative AI models (ICO, 2024[92]). These include the requirements developers must meet in terms of complying with data subject rights and the accuracy principle in the UK GDPR, as well as the lawful basis for web scraping to train generative AI models.

The **United States** Federal Trade Commission (FTC) has provided guidance on the use of algorithms in automated decision-making. The blog post titled “Using Artificial Intelligence and Algorithms” highlighted the potential benefits and risks presented by increasingly sophisticated technologies, particularly in the domain of AI and healthcare (FTC, 2020[93]). Another blog post on Truth, Fairness, and Equity in AI focuses on how existing US laws can prevent the use of AI that is biased or unfair (FTC, 2021[94]). FTC staff have also warned businesses to avoid using automated tools that have biased or discriminatory impacts (FTC, 2023[95]). On 25 April 2023, the FTC and officials from three other federal agencies (Consumer Financial Protection Bureau, Justice Department’s Civil Rights Division, and Equal Employment Opportunity Commission) also signed a Joint Statement on enforcement efforts against discrimination and bias in automated systems (FTC, 2023[96]).

In **Singapore**, the Personal Data Protection Commission issued Advisory Guidelines on the use of personal data in AI Recommendation and Decision Systems. The Advisory Guidelines seek to provide i) organisations with more clarity on the use of personal data to train or develop AI to support their efforts to implement AI; ii) guidance on information to be provided to consumers when seeking consent; iii) guidance to third-party developers of AI systems who may occupy the role of data intermediaries; iv) guidance on best practices to support businesses in their compliance with the Personal Data Protection Act (PDPC, 2022[97]).

PEA enforcement actions in AI, including generative AI

PEA's have taken enforcement actions at the intersection of AI and privacy, including in response to generative AI. Such efforts have largely been focused on OpenAI as the provider of ChatGPT.

The Federal Office of the Privacy Commissioner (OPC) of **Canada** announced on 4 April 2023, that it has launched an investigation into ChatGPT following a complaint that the service is processing personal data without consent. On 25 May, the OPC announced that it will investigate ChatGPT jointly with the provincial privacy authorities of British Columbia, Quebec, and Alberta, expanding the investigation to also look into whether OpenAI has respected obligations related to openness and transparency, access, accuracy, and accountability, as well as purpose limitation.

The **Italian** PEA (Garante) issued an emergency order on 30 March 2023 to block OpenAI from processing personal data of people in Italy. The Garante laid out several potential violations of provisions of the GDPR, including lawfulness, transparency, rights of the data subject, processing personal data of children, and data protection by design and by default. It lifted the prohibition a month later, after OpenAI announced changes as required by the Garante.

Japan's Personal Information Protection Commission (PPC) published a warning issued to OpenAI on 1 June 2023 which highlighted that OpenAI should not collect sensitive personal data from users of ChatGPT or other persons without obtaining consent, and it should give notice in Japanese about the purpose for which it collects personal data from users and non-users.

The Personal Information Protection Commission of **Korea** (PIPC) announced on 27 July 2023 that it imposed an administrative fine of 3.6 million KRW (approximately USD 3,000) against OpenAI for failure to notify a data breach in relation to its payment procedure. At the same time, the PIPC issued a list of instances of non-compliance with the country's Personal Information Protection Act (PIPA) related to transparency, lawful grounds for processing (absence of consent), lack of clarity related to the controller-processor relationship, and issues related to the absence of parental consent for children younger than 14 years of age. The PIPC gave OpenAI until 15 September 2023, to bring the processing of personal data into compliance.

In the **United Kingdom**, in May 2022 the ICO fined United-based Clearview AI Inc. GBP 7,552,800 for using images of people in the United Kingdom, and elsewhere, that were collected from the web and social media to create a global online database that could be used for facial recognition. The ICO also issued an enforcement notice, ordering the company to stop obtaining and using the personal data of United Kingdom residents that is publicly available on the internet, and to delete the data of United Kingdom residents from its systems. Clearview AI appealed both notices issued by ICO. On 17 October 2023, the First Tier Tribunal (FTT) supported the ICO's view that Clearview AI processed personal information related to monitoring individuals' behavior through the collection of billions of facial images. These images were then provided for access and analysis using AI to foreign subscribers. However, the FTT ruled that the ICO lacked enforcement powers against Clearview AI because its clients were limited to foreign law enforcement and government agencies discharging criminal law or national security functions, falling outside the scope of the GDPR and UK GDPR. In November 2023, the United Kingdom Information Commissioner sought permission to appeal the Clearview AI judgment of the First Tier Tribunal (Information Rights). The ICO issued a preliminary enforcement notice in October 2023 against Snap Inc., the parent company of Snapchat, over its potential failure to properly assess the privacy risks posed by its generative AI chatbot 'My AI', including risks to children aged 13-17.

In the **United States**, the FTC has brought multiple enforcement actions to protect consumers from abuses involving AI, including a case against Rite Aid, a large retail pharmacy chain, that employed a facial recognition technology which purported to match photos of customers inside their stores with photos appearing in a database of shoplifters and troublemakers. The FTC alleged that in areas where the plurality

of the population was Black or Asian, the facial recognition technology was significantly more likely to result in false positive facial recognition matches, resulting in innocent people being confronted by store employees and asked to leave the stores. The FTC has also brought two cases against Amazon concerning the company's Alexa App and its Ring cameras. In the Amazon Alexa case, the FTC charged Amazon with, among other things, violating the Children's Online Privacy Protection Act and obtained a final order requiring Amazon to delete inactive child accounts and certain voice recordings and geolocation information and prohibiting the company from using such data to train its algorithms. In the Amazon Ring case, the final order requires Amazon to delete any customer videos and face embeddings – data collected from an individual's face – that it obtained prior to 2018 and to delete any work products derived from these videos. The FTC's other enforcement actions include cases focusing on claims regarding the use of algorithms and AI in investment recommendation programs (FTC, 2022[98]) (FTC, 2021[99]).

The **Brazilian** PEA announced on 27 July 2023 that it started an investigation into how ChatGPT is complying with the Lei Geral de Proteção de Dados (LGPD) after receiving a complaint, and after reports in the media arguing that the service as provided is not compliant with the country's data protection law.

In the aftermath of the Italian order, the **European Data Protection Board** (EDPB) created a task force on 13 April 2023 to "foster co-operation and exchange information" in relation to handling complaints and investigations into OpenAI and ChatGPT at European Union level. On May 2024, the EDPB released a report detailing the efforts and preliminary views on certain aspects of the investigation. The preliminary views address the compliance of ChatGPT with key GDPR principles, including lawfulness, fairness, transparency, data accuracy, and the rights of data subjects (EDPB, 2024[100]).

The Ibero-American Network of Data Protection (RIPD), reuniting supervisory authorities from 21 Spanish- and Portuguese-speaking countries in Latin America and Europe, announced on 8 May 2023 that it initiated a coordinated action in relation to ChatGPT.

This preliminary stock taking of investigations into how generative AI service providers are complying with data protection law in jurisdictions around the world reveals significant commonalities among their legal obligations and how they are applicable to the processing of personal data when it comes to generative AI. It also highlights the importance of international co-operation in the enforcement of privacy laws, which is the focus of separate work by the OECD in the context of the revision of the Recommendation on Cross-Border Co-operation in the Enforcement of Laws Protecting Privacy, originally adopted in 2007 (OECD, 2007[6])

4 Conclusion

This report provides an overview of priority areas where the AI and privacy communities can collaborate and strengthen their synergies within the OECD and beyond, and help develop resources and tools for trustworthy AI systems that respect privacy. It captures policy opportunities and challenges of joint relevance to these communities, areas of complementarity and potential gaps.

Recently, the OECD has decided to play its part, leveraging its unique co-operation infrastructure and convening power across borders, sectors and areas of expertise to support and foster a positive and proactive message about AI and privacy. Established in early 2024, the OECD.AI Expert Group on AI, Data, and Privacy advances analysis into these areas to strengthen connections between the AI and privacy communities. This includes considering the creation of common guidance and recommendations to facilitate global interoperability across AI and privacy governance. In collaboration with relevant experts at the OECD and beyond, the Expert Group could consider working on sector-specific issues at the intersection of AI and privacy, such as in health, employment, or finance.

In the medium term, these co-operation efforts could help assess whether the AI and Privacy Recommendations need to be updated to reflect the synergies between the AI and privacy communities. The work of the newly created expert group could also lead to the development of practical resources for AI actors and privacy and data protection regulators.

International co-operation on AI and privacy should aim for the long-term interoperability of legal, technical, and operational frameworks applying to AI and privacy. This will allow policy and decision-makers to leverage commonalities, complementarities, and elements of convergence in their respective policy frameworks, or, conversely, to identify the stumbling blocks that could hinder the development of common positions or co-operation.

References

Abid, A. (2021), *"Persistent Anti-Muslim Bias in Large Language Models"*, http://arxiv.org/abs/2101.05783. [34]

AEPD (2023), *Artificial Intelligence: Transparency*. [89]

AEPD (2021), *Audit Requirements for Personal Data Processing Activities involving AI*. [88]

AEPD (2020), *Guide on AI-based data processing*. [72]

Ahmed Salem, Y. (2018), *ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models*, https://arxiv.org/abs/1806.01246. [30]

Barros Vale, S. (2022), *GDPR AND THE AI ACT INTERPLAY: LESSONS FROM FPF'S ADM CASE-LAW REPORT*, https://fpf.org/blog/gdpr-and-the-ai-act-interplay-lessons-from-fpfs-adm-case-law-report/. [38]

BCG (2023), *What's Dividing the C-Suite on Generative AI?*, https://www.bcg.com/publications/2023/c-suite-genai-concerns-challenges?utm_source=linkedin&utm_medium=social&utm_campaign=digital-transformation&utm_description=paid&utm_topic=ai&utm_geo=global. [105]

Bommasani, R. (2021), *On the opportunities and risks of foundation models*, https://arxiv.org/abs/2108.07258. [73]

CJEU (2022), *Ligue des droits humains ASBL v Conseil des ministres(ECLI:EU:C:2022:491)*, https://curia.europa.eu/juris/document/document.jsf;jsessionid=64DD86ED8C10F3FF696478DEE8FF540A?text=&docid=261282&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=3035820 . [69]

Clifford, D. and J. Ausloos (2018), "Data Protection and the Role of Fairness", *Yearbook of European Law*, Vol. 37, pp. 130-187, https://doi.org/10.1093/yel/yey004. [55]

CNIL (2024), *Les fiches pratiques IA*, https://www.cnil.fr/fr/les-fiches-pratiques-ia. [87]

CNIL (2023), *AI how-to sheets*. [63]

CNIL (2023), *Artificial intelligence: the action plan of the CNIL*. [86]

CNIL (2023), *Data, footprints and freedoms: discover the CNIL's new Innovation & Foresight report*, https://www.cnil.fr/en/data-footprints-and-freedoms-discover-cnils-new-innovation-foresight-report. [42]

CNIL (2022), *Chacun chez soi et les données seront bien gardées : l'apprentissage fédéré*, https://linc.cnil.fr/chacun-chez-soi-et-les-donnees-seront-bien-gardees-lapprentissage-federe. [111]

CNIL (2017), *HOW CAN HUMANS KEEP THE UPPER HAND?*, https://www.cnil.fr/sites/cnil/files/atoms/files/cnil_rapport_ai_gb_web.pdf. [58]

Comiter, M. (2019), "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It", *Belfer Center Paper*. [76]

Datatilsynet (2018), *Artificial intelligence and privacy*, https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf. [57]

Datatilsynet (2018), *Artificial intelligence and privacy*. [110]

Doshi-Velez, F. (2017), "Accountability of AI Under the Law: The Role of Explanation", *Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein* 10, https://dash.harvard.edu/handle/1/34372584. [112]

EDPB (2024), *Report of the work undertaken by the ChatGPT Taskforce*, https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf. [100]

European Commission and US TTC (2023), *EU-U.S. Terminology and Taxonomy for Artificial Intelligence*. [5]

European Parliament (2024), *Artificial Intelligence Act: MEPs adopt landmark law*, European Parliament, https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law. [2]

European Union (2016), *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, https://eur-lex.europa.eu/eli/reg/2016/679/oj. [43]

FTC (2023), *Joint Statement on enforcement efforts against discrimination and bias in automated systems*, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf. [96]

FTC (2023), *Keep your AI claims in check*, https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check. [95]

FTC (2022), *FTC Takes Action Against Company Formerly Known as Weight Watchers for Illegally Collecting Kids' Sensitive Health Data*, https://www.ftc.gov/news-events/news/press-releases/2022/03/ftc-takes-action-against-company-formerly-known-weight-watchers-illegally-collecting-kids-sensitive. [98]

FTC (2022), *In the Matter of Passport Auto Group(File No. 2023199)*, https://www.ftc.gov/system/files/ftc_gov/pdf/joint-statement-of-chair-lina-m.-khan-commissioner-rebecca-kelly-slaughter-and-commissioner-alvaro-m.-bedoya-in-the-matter-of-passport-auto-group.pdf (accessed on 25 October 2023). [50]

FTC (2021), *Aiming for truth, fairness, and equity in your company's use of AI*, https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai. [94]

FTC (2021), *FTC Finalizes Settlement with Photo App Developer Related to Misuse of Facial Recognition Technology*, https://www.ftc.gov/news-events/news/press-releases/2021/05/ftc-finalizes-settlement-photo-app-developer-related-misuse-facial-recognition-technology. [99]

FTC (2020), *Using Artificial Intelligence and Algorithms*, https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms. [93]

G7 (2023), *Roundtable of G7 Data Protection and Privacy Authorities Statement on Generative AI*, https://www.ppc.go.jp/files/pdf/G7roundtable_202306_statement.pdf. [35]

Garante per la protezione dei dati personali (2024), *ChatGPT: Garante privacy, notificato a OpenAI l'atto di contestazione per le violazioni alla normativa privacy*, https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9978020#english. [36]

Global Privacy Assembly (2023), *Joint statement on data scraping and the protection of privacy*, https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf. [83]

Global Privacy Assembly (2023), *Resolution on Artificial Intelligence and Employment*, https://www.edps.europa.eu/system/files/2023-10/1.-resolution-on-ai-and-employment-en.pdf. [84]

Global Privacy Assembly (2023), *Resolution on Generative Artificial Intelligence Systems*, https://www.edps.europa.eu/system/files/2023-10/edps-gpa-resolution-on-generative-ai-systems_en.pdf. [82]

GPA (2023), *Resolution on Generative Artificial Intelligence Systems*. [15]

GPA's International Enforcement Cooperation Working Group (2023), *Joint statement on data scraping and the protection of privacy*, https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf. [28]

Grazia, M. (2017), "Use of the Charter of Fundamental Rights by National Data Protection Authorities and the EDPS", *RSCAS Research Project Reports, Centre for Judicial Cooperation*. [39]

Hannah Brown, K. (2022), "What Does it Mean for a Language Model to Preserve Privacy?", *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, https://doi.org/10.1145/3531146.3534642. [31]

Harvard, J. (2024), *Differential privacy*, https://privacytools.seas.harvard.edu/differential-privacy. [22]

Ibero-American Data Protection Network (2020), *General Recommendations for the Processing of Personal Data in Artificial Intelligence*. [90]

Ibero-American Data Protection Network (2020), *Specific Guidelines for Compliance with the Principles and Rights that Govern the Protection of Personal Data in Artificial Intelligence Projects.*. [91]

ICO (2024), *Consultation series on generative AI and data protection*, https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-consultation-series-on-generative-ai-and-data-protection/. [92]

ICO (2023), *Guidance on AI and data protection*. [108]

ICO (2023), *Guidance on AI and data protection*. [54]

ICO (2023), *How should we assess security and data minimisation in AI?*. [59]

ICO (2020), *Explaining decisions made with AI*. [67]

Kleinberg, J. and M. Raghavan (2021), "Algorithmic monoculture and social welfare", *Proceedings of the National Academy of Sciences*, Vol. 118/22, https://doi.org/10.1073/pnas.2018340118. [74]

Kovacs, E. (2023), *ChatGPT Data Breach Confirmed as Security Firm Warns of Vulnerable Component* [75]

Exploitation, https://www.securityweek.com/chatgpt-data-breach-confirmed-as-security-firm-warns-of-vulnerable-component-exploitation/.

Liu, A. (2022), *Threats, attacks and defenses to federated leanring: issues, taxonomy and perspectives*, https://doi.org/10.1186/s42400-021-00105-6. [113]

Liwei Song, R. (2019), *Privacy Risks of Securing Machine Learning Models against Adversarial Examples*, https://arxiv.org/abs/1905.10291. [77]

Lorenz, P., K. Perset and J. Berryhill (2023), "Initial policy considerations for generative artificial intelligence"*, OECD Artificial Intelligence Papers*, No. 1, OECD Publishing, Paris, https://doi.org/10.1787/fae2d1e6-en. [12]

Malgieri, G. (2020), "The concept of fairness in the GDPR", *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, https://doi.org/10.1145/3351095.3372868. [53]

Margaret Mitchell, S. (2019), *Model cards for model reporting*, https://dl.acm.org/doi/10.1145/3287560.3287596. [66]

Maxwell, W. and B. Dumas (2023), "Meaningful XAI based on user-centric design methodology: Combining legal and human-computer interaction (HCI) approaches to achieve meaningful algorithmic explainability.", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.4520754. [71]

MIT (2022), *Collaborative machine learning that preserves privacy*, https://news.mit.edu/2022/collaborative-machine-learning-privacy-0907. [21]

Mulligan, D. et al. (2019), "This Thing Called Fairness", *Proceedings of the ACM on Human-Computer Interaction*, Vol. 3/CSCW, pp. 1-36, https://doi.org/10.1145/3359221. [56]

Mulligan, D. et al. (2021), "Confidential Computing - a brave new world", *2021 International Symposium on Secure and Private Execution Environment Design (SEED)*, https://doi.org/10.1109/SEED51797.2021.00025. [20]

NIST (2024), *Adversarial Machine Learning*. [80]

NIST (2023), *Artificial Intelligence Risk Management*. [79]

NIST (2021), *Four Principles of Explainable Artificial Intelligence*. [68]

O'Brien, C. (2020), *Why Intel believes confidential computing will boost AI and machine learning*, https://venturebeat.com/ai/why-intel-believes-confidential-computing-will-boost-ai-and-machine-learning/. [19]

OECD (2024), *Digital Economy Outlook Chapter 2: The Future of Artificial Intelligence*, OECD Publishing, https://www.oecd-ilibrary.org/sites/a1689dc5-en/1/3/2/index.html?itemId=/content/publication/a1689dc5-en&_csp_=5cbbea11094afe4b75c96b4a3ec0bcd2&itemIGO=oecd&itemContentType=book. [1]

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. [61]

OECD (2023), *"Explanatory memoranda of the OECD Privacy Guidelines", OECD Digital Economy Papers, No. 360, OECD Publishing, Paris*, https://doi.org/10.1787/ea4e9759-en. [60]

OECD (2023), "Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI", *OECD Digital Economy Papers*, No. 349, OECD Publishing, Paris, https://doi.org/10.1787/2448f04b-en. [14]

OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *OECD Digital Economy Papers*, No. 352, OECD Publishing, Paris, https://doi.org/10.1787/13d38f92-en. [13]

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. [8]

OECD (2023), "G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI", Vol. OECD Publishing, Paris, https://doi.org/10.1787/bf3c0c60-en. [3]

OECD (2023), *Generative artificial intelligence in finance*, https://doi.org/10.1787/ac7149cc-en. [49]

OECD (2023), *Initial policy considerations for generative artificial intelligence*, https://doi.org/10.1787/fae2d1e6-en. [115]

OECD (2023), *Moving forward on data free flow with trust: New evidence and analysis of business experiences*, OECD Publishing, Paris, https://doi.org/10.1787/1afab147-en. [40]

OECD (2023), *OECD Global Forum on Digital Security for Prosperity, session on Security risks in artificial intelligence*, https://www.oecd.org/digital/global-forum-digital-security/GFDSP-2023-agenda.pdf. [18]

OECD (2023), *OECD Privacy Guidelines Implementation Guidance: Foreword and Chapter on Accountability*, https://one.oecd.org/official-document/DSTI/CDEP/DGP(2022)8/FINAL/en. [44]

OECD (2023), *Report on the Implementation of the OECD Privacy Guidelines*, https://www.oecd-ilibrary.org/docserver/cf87ae8f-en.pdf?expires=1714657586&id=id&accname=ocid84004878&checksum=88C95E8CD9F60E5B2490815D12CC68FD. [16]

OECD (2023), "Summary of the OECD-MIT virtual roundtable on the future of artificial intelligence (AI)", OECD, Paris, https://wp.oecd.ai/app/uploads/2023/03/OECD-MIT-Workshop-1.pdf. [23]

OECD (2022), *Companion Document to the OECD Recommendation on Children in the Digital Environment*, https://www.oecd-ilibrary.org/science-and-technology/companion-document-to-the-oecd-recommendation-on-children-in-the-digital-environment_a2ebec7c-en. [48]

OECD (2022), *Companion Document to the OECD Recommendation on Children in the Digital Environment*, https://www.oecd-ilibrary.org/science-and-technology/companion-document-to-the-oecd-recommendation-on-children-in-the-digital-environment_a2ebec7c-en. [123]

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. [17]

OECD (2022), *Summary of the OECD-MIT virtual roundtable on the future of Artificial Intelligence (AI)*, https://wp.oecd.ai/app/uploads/2023/03/OECD-MIT-Workshop-1.pdf. [102]

OECD (2021), *Children in the Digital Environment: Revised Typology of Risks*, https://www.oecd-ilibrary.org/docserver/9b8f222e-en.pdf?expires=1712825182&id=id&accname=guest&checksum=A12040CCC20943E33499684B1C2A1187. [47]

OECD (2021), *Children in the Digital Environment: Revised Typology of Risks*, https://www.oecd-ilibrary.org/docserver/a2ebec7c-en.pdf?expires=1653053778&id=id&accname=guest&checksum=AD7D1497BD71DD3BA15ABD5B1557BF8F. [121]

OECD (2021), *Guidelines for Digital Service Providers*, https://www.oecd.org/mcm/OECD%20Guidelines%20for%20Digital%20Service%20Providers.pdf. [46]

OECD (2021), *Guidelines for Digital Service Providers*, https://one.oecd.org/document/C/MIN(2021)7/ADD1/en/pdf. [122]

OECD (2021), *Recommendation of the Council on Children in the Digital Environment*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389. [45]

OECD (2021), *Recommendation of the Council on Children in the Digital Environment*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389%20. [118]

OECD (2021), *RECOMMENDATION OF THE COUNCIL ON ENHANCING ACCESS TO AND SHARING OF DATA*, https://www.oecd.org/mcm/Recommendation-of-the-Council-on-Enhancing-Access-to-and-Sharing-of-Data_EN.pdf. [7]

OECD (2021), *REPORT ON THE IMPLEMENTATION OF THE RECOMMENDATION OF THE COUNCIL CONCERNING GUIDELINES GOVERNING THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA*, https://one.oecd.org/document/C(2021)42/en/pdf. [9]

OECD (2020), *Protection of Children Online: An Overview of Recent Developments in Legal Frameworks and Policies*, https://www.oecd-ilibrary.org/docserver/9e0e49a9-en.pdf?expires=1653410836&id=id&accname=ocid84004878&checksum=4A109A32D62EB1DA07305E7CEE85747D. [120]

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. [11]

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [106]

OECD (2017), *Protection of Children Online: Preliminary Country Survey Findings and Proposal For Next Steps*, https://one.oecd.org/document/DSTI/CDEP/SPDE(2017)3/en/pdf. [119]

OECD (2011), *The Protection of Children Online: Risks Faced by Children Online and Policies to Protect Them*, https://www.oecd-ilibrary.org/science-and-technology/the-protection-of-children-online_5kgcjf71pl28-en. [117]

OECD (2008), *Declaration for the Future of the Internet Economy (The Seoul Declaration)*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0366. [116]

OECD (2007), *RECOMMENDATION OF THE COUNCIL ON CROSS-BORDER CO-OPERATION IN THE ENFORCEMENT OF LAWS PROTECTING PRIVACY*, https://one.oecd.org/document/C(2007)67/FINAL/en/pdf. [6]

OECD (1980), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188. [114]

OECD (Updated 2023), *Definition of an AI System*. [41]

OECD AI (2023), *About OECD.AI*, https://oecd.ai/en/about/the-context. [10]

OECD.AI (2024), *Catalogue of Tools & Metrics for Trustworthy AI*, https://oecd.ai/en/catalogue/tools?terms=fairness&page=1. [64]

Office of the Privacy Commissioner of Canada (2023), *Canadian privacy regulators launch principles for responsible development and use of generative AI*. [85]

Office of the Privacy Commissioner of Canada (2023), *priv.gc.ca*. [37]

OPC (2022), *When what is old is new again – the reality of synthetic data*, OPC Privacy Tech-know blog, https://priv.gc.ca/en/blog/20221012/?id=7777-6-493564 (accessed on 1 November 2023). [25]

PDPC (2022), *Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems*, https://www.pdpc.gov.sg/guidelines-and-consultation/2024/02/advisory-guidelines-on-use-of-personal-data-in-ai-recommendation-and-decision-systems. [97]

Pew Research Center (2023), *How Americans View Data Privacy*, https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/. [103]

Reza Shokri, M. (2019), *On the Privacy Risks of Model Explanations*, https://arxiv.org/abs/1907.00164. [78]

Robin Staab, M. (2023), "Beyond Memorization: Violating Privacy Via Inference with Large Language Models", https://arxiv.org/abs/2310.07298. [29]

Robinson, L., K. Kizawa and E. Ronchi (2021), "Interoperability of privacy and data protection frameworks", *OECD Going Digital Toolkit Notes*, No. 21, OECD Publishing, Paris, https://doi.org/10.1787/64923d53-en. [101]

Rotenberg, M. (2021), "Artificial Intelligence and Democratic Values: The Role of Data Protection", *European Data Protection Law Review*, Vol. 7/4, pp. 496-501, https://doi.org/10.21552/edpl/2021/4/6. [70]

Shumailov, A. (2023), *The Curse of Recursion: Training on Generated Data Makes Models Forget*, https://arxiv.org/abs/2305.17493. [26]

Singapore Data Protection Commission (2024), *ADVISORY GUIDELINES ON USE OF PERSONAL DATA IN AI RECOMMENDATION AND DECISION SYSTEMS*. [65]

Solove, D. (2024), "Artificial Intelligence and Privacy", *77 FLORIDA LAW REVIEW (forthcoming)*, pp. 5-8. [32]

Stadler, T., B. Oprisanu and C. Troncoso (2020), "Synthetic data – Anonymisation Groundhog Day", *Proceedings of the 31st USENIX Security Symposium, Security 2022*, pp. 1451-1468, https://doi.org/10.48550/arxiv.2011.07018. [24]

Tabassi, E. (2023), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, https://doi.org/10.6028/nist.ai.100-1. [51]

Tarun, A. (2023), *https://arxiv.org/pdf/2111.08947.pdf*, https://arxiv.org/pdf/2111.08947.pdf. [27]

The White House (2023), *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/. [107]

The White House (2023), *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/. [4]

the White House (2023), *FACT SHEET: Biden-□Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/. [104]

Wachter, S. (2022), "THE THEORY OF ARTIFICIAL IMMUTABILITY: PROTECTING ALGORITHMIC GROUPS UNDER ANTI-DISCRIMINATION LAW", https://doi.org/10.48550/arXiv.2205.01166 (accessed on 25 October 2023). [52]

Weidinger, L. (2022), *"Taxonomy of Risks Posed by Language Models"*, https://doi.org/10.1145/3531146.3533088. [33]

Widder, D. (2023), *Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI*, http://dx.doi.org/10.2139/ssrn.4. [62]

Widder, D. (2023), "Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI", http://dx.doi.org/10.2139/ssrn.4. [109]

Yeung, K., A. Howes and G. Pogrebna (2020), *AI Governance by Human Rights–Centered Design, Deliberation, and Oversight*, Oxford University Press, https://doi.org/10.1093/oxfordhb/9780190067397.013.5. [81]

Notes

[1] According to the OECD Privacy Guidelines, Privacy Enforcement Authority means "any public body, as determined by each Member country, that is responsible for enforcing laws protecting privacy, and that has powers to conduct investigations or pursue enforcement proceedings" (OECD, 1980[114]).

[2] "Synthetic data is generated from data/processes and a model that is trained to reproduce the characteristics and structure of the original data aiming for similar distribution. The degree to which synthetic data is an accurate proxy for the original data is a measure of the utility of the method and the model." (European Commission and US TTC, 2023[5]).