48th Closed Session of the Global Privacy Assembly

November 2026

Resolution on Privacy-protective Agentic AI

This Resolution is submitted by the Working Group on Ethics and Data Protection in Artificial Intelligence (AIWG).

SPONSORS

- Office of the Privacy Commissioner of Canada
- Autoritat Catalana de Protecció de Dades
- Office of the Privacy Commissioner for Personal Data, Hong Kong, China
- Personal Information Protection Commission, Korea
- Agencia Española de Protección de Datos
- Federal Data Protection and Information Commissioner of Switzerland
- Information Commission's Office, United Kingdom

CO-SPONSORS

- Agencia de Acceso a la Información Pública (Argentina)
- Office of the Information and Privacy Commissioner of British Columbia (Canada)
- European Data Protection Supervisor
- Commission Nationale de l'Informatique et des Libertés, France
- Federal Commissioner for Data Protection and Freedom of Information (BfDI), Germany
- Rhineland Palatinate Data Protection Commissioner, Germany
- Israeli Privacy Protection Authority
- Isle of Man Information Commissioner
- Office of the Information and Privacy Commissioner of Ontario (Canada)

The 48th Global Privacy Assembly 2026:

Acknowledging the emergence and growing adoption of artificial intelligence (AI) agents and agentic AI systems;

Understanding, based on a definition from the OECD, 'AI agent' to mean a software entity that perceives and acts upon its environment with a degree of autonomy, using tools as needed to achieve specific goals and adapt to changing inputs and contexts;

Understanding, again based on an OECD definition, an 'agentic AI system' to refer to a system composed of one or more AI agents that can break down tasks, collaborate where applicable, and pursue complex objectives autonomously over extended periods, and designed to operate in open-ended or less predictable physical or virtual environments and to function with minimal human supervision;

Recognizing that the foundational component underpinning most current agentic AI systems is a Large Language Model (LLM), that the data protection challenges inherent to such models will also apply to agentic AI systems, and that as this Resolution is based on this recognition some considerations may need to evolve if other bases for agentic AI systems become common;

Recognizing as well that individual data protection authorities may further refine these understandings, such as by defining agentic AI systems to play an orchestration (rather than coordination) role for AI agents;

Limiting the scope of this resolution to those systems and situations in which an agentic AI system processes the personal data of one or more individuals, including where such systems are used to make automated decisions about individuals or take actions affecting them;

Noting as well that while tools are increasingly being introduced to allow individuals to create AI agents or utilize agentic AI systems in a personal capacity, the scope of the current resolution will be limited to the use, deployment or distribution of AI agents and agentic AI systems within and by public and private organizations;

Being aware of the potential for a human-centric approach to AI to bring important economic, societal and public interest benefits, including by growing prosperity and addressing pressing global challenges;

Emphasizing the importance of ensuring that the use of AI systems respects human rights including non-discrimination, is fair and lawful, and prevents bias;

Recognizing that the adoption of agentic AI systems could increase the potential for significant adverse effects associated with automated decision-making processes, particularly should such systems reduce human oversight, increase automation bias, or limit the opportunity for impacted individuals to meaningfully challenge decisions;

Emphasizing that organizations remain accountable for personal data processed through agentic AI systems (including where that processing is initiated autonomously by AI agents or results from their use of tools, skills, other agents or connected services) and that organizations should not assume that such systems will recognize or comply with organizational rules, contextual expectations or limitations;

Recognizing as well that with appropriate and demonstrable safeguards in place, agentic AI systems could be used to strengthen an organization's privacy practices by making processes more efficient, consistent, and auditable, supporting the execution of key tasks (such as the processing of data subject access requests), or reducing the frequency with which sensitive data is made available to individuals within the organization, while also ensuring that legally or contextually sensitive decisions remain subject to human judgment;

Recognizing the importance of understanding the characteristics and differentiations associated with new technologies that could lead to novel challenges for privacy and data protection;

Acknowledging that while agentic AI systems raise new and novel privacy concerns such systems will generally also raise many of the same issues associated with other forms of AI, and thus that existing Global Privacy Assembly Resolutions and Statements on the topic remain relevant, including the Declaration on Ethics and Data Protection in Artificial Intelligence (40th GPA), the Resolution on Accountability in the Development and Use of Artificial Intelligence (42nd GPA), the Resolution on AI and Employment and the Resolution on Generative AI Systems (45th GPA), and the Resolution on the Collection, Use and Disclosure of Personal Data to Pre-Train, Train and Fine-Tune AI Models, and the Resolution on Meaningful Human Oversight of Decisions Involving AI Systems (47th GPA);

Highlighting as well the relevant work undertaken by other networks of data protection authorities, including the Association Francophone des Autorités de Protection des Données Personnelles (AFAPDP)'s Resolution on the Ethical Governance of Artificial Intelligence, the Roundtable of G7 Data Protection Authorities Statements on Generative AI, AI and Children, and the Role of Data Protection Authorities in Fostering Trustworthy AI, and the Joint Statement on Building Trustworthy Data Governance Frameworks to Encourage Development of Innovative and Privacy-Protective AI signed by 20 data protection authorities;

Emphasizing that in addition to the privacy considerations set out in the documents above, AI agents and agentic AI systems can create, exacerbate, or highlight privacy challenges, including with respect to:

- **Dynamic tool use**: The AI agents within an agentic AI system, or the tools, skills, plugins, and other components that they use or call, may be developed, supplied or controlled by a series of different organizations and operators all of which could have differing, changing or inconsistent privacy practices, contractual terms, and technical safeguards. Further, in addition to the challenge in a traditional AI system that a user may not know in advance the logic that will be applied to reach a conclusion, in an agentic AI system the user may also not know at the outset which components may be called or used and thus not know how personal data will flow between parties – a consideration exacerbated by the brittleness of the approaches taken by agentic AI systems (i.e. small changes in a prompt may have significant impacts on them). This creates challenges for organizations with respect to maintaining accountability over data processing, managing any cross-border transfer of data, providing sufficient transparency about data practices, and ensuring individuals can effectively exercise their privacy rights. The use of multiple tools also creates the potential for cascading inaccuracies, in which a single incorrect output from a tool can negatively impact the accuracy of all other tools.

- **Memory**: AI agents may maintain a 'working memory', which can include information about prior tasks (for instance, to aid the agent in identifying the most efficient or effective steps to complete similar tasks in the future). If such memory includes personal data, risks may arise related to inaccurate or excessive data, retention beyond the period necessary for the purpose, or manipulation of the agent that impacts the confidentiality, integrity or availability of that data. Should an AI agent be used across multiple tasks by multiple users, the risk of stored personal data being inadvertently disclosed increases.
- **Data access and delegated authority**: AI agents may have access to multiple sources of data, interfaces, accounts, or systems in the environment in which they operate. These may be internal or external, or accessible through the use of tools. This requires consideration of whether an agent is accessing the minimum data and capabilities required to complete a task; whether it is unnecessarily processing sensitive data; whether personal data made available for one task may be combined with other data or reused for a different or incompatible purpose; and, particularly in the case of external data, whether the data made available is sufficiently complete, accurate and reliable for its purpose. There is also a risk that agents can be manipulated into inappropriately disclosing or giving unauthorized parties access to the data sources to which they have access, modifying records, changing permissions, communicating with unauthorized parties or taking other consequential and unintended or inappropriate actions, each of which is more likely to occur if components are insecure, compromised, obsolete or inappropriately permissioned. Finally, the delegated authority given to an agent may mean that it is difficult to distinguish actions initiated directly by a user from those taken autonomously by an AI agent.
- **Security**: Agentic AI systems' tool use and data access creates an increased attack surface, which in turn introduces or amplifies security vulnerabilities. An agentic AI system may, for example, incorporate data from an external source or the output of a third-party tool into the context for the next step of a process, either of which creates the potential for data poisoning and prompt-injection attacks. Agentic AI systems with broad access to the personal data held by an organization could be particularly attractive targets for such attacks. Agentic AI systems' capacity to carry out actions – for example, through tool use – also gives rise to an increased security risks, as a compromised agent with privileges within an organization could rapidly expand the scope and impact of a breach. Agentic AI systems could also be used to enhance the speed or effectiveness of cyberattacks.
- **Human Oversight and Review**: A defining characteristic of an agentic AI system is its ability to impact its environment or take action semi- or fully-autonomously. Thus, to achieve greatest perceived benefit or efficiency, organizations may seek to limit the number of times an agentic AI system pauses operation to seek human input such as approvals. This, in turn, may reduce the ability to exercise meaningful human oversight of AI-based decision-making systems, or to allow individuals to effectively request human review of an automated decision. The semi- or fully-autonomous nature of agents may also mean that the process by

which an agent has arrived at a determination or decision is difficult to explain, or the validity and reliability of the determination or decision is difficult to confirm.

- **Unauthorized or ungoverned deployment:** As with other AI-based tools, individuals within organizations may believe that their productivity will be enhanced through the adoption of agentic AI systems. The growing accessibility of such tools may also encourage individuals to deploy or use them for organizational purposes without appropriate approval or oversight. However, unlike more traditional AI systems into which an individual would need to proactively upload data, the configurability and ability of agentic AI systems to connect to organizational data, tools and systems may expose significant amounts of personal data to AI agents with few or no controls. Such use may bypass organizational policies, data inventories, access controls, contractual safeguards and incident-response arrangements, and may expose personal data or internal systems to unapproved agents, tools or external services.

Recognizing that organizations have privacy obligations when using or deploying agentic AI systems, they should implement risk-based measures to meet these obligations and to mitigate the negative impacts of the considerations described above. Examples of such measures include:

- **Establishing appropriate review and monitoring mechanisms** throughout the lifecycle of an agentic AI system from preparation and deployment to use and eventual decommissioning, which could include data protection impact assessments (DPIAs) to ensure that privacy risks are identified and mitigated prior to deployment, technical auditing of agentic AI systems during procurement or in advance of deployment, and periodic assessment of the validity, accuracy, reliability and suitability of agentic AI systems. These measures should be revisited periodically as well as following material changes, incidents or the identification of emerging vulnerabilities. These measures could also be supported by specialized AI agents.
- **Determining appropriate limits for the uses of agentic AI systems**, proportionate to the nature of the processing, the sensitivity of data, and the potential impact on individuals. In some cases, this could include limiting actions taken or decisions made by agentic AI systems to those that are revocable or reversible with no or minimal harm to individuals.
- **Establishing policies governing agent access to, use, retention, deletion and disclosure of data assets**. This could include maintaining inventories of what data an agent is able to access; limiting the circumstances under which an agent can access, use or disclose sensitive data; clearly defining the purposes for which personal data may be accessed, processed or used by agentic AI systems, minimizing the amount of personal data that an agent can access for those purposes, and implementing safeguards to prevent the use of such data beyond those purposes; pseudonymizing data before it can be accessed by an AI agent; ensuring the accuracy of data accessible to AI agents; putting in place appropriate validation and control mechanisms with respect to the tools accessible to the AI agents and controlling the amount and/or nature of data that can be disclosed to those tools; or, ensuring that personal data processed by the AI agent or agentic AI system is not retained longer than is necessary. This could also include secure deployment architectures appropriate to the nature and risk of the

processing, including proportionate measures to isolate or contain agentic AI systems' operating environments, restricting unnecessary network exposure, and preventing failures or compromise from affecting other systems or data. Finally, organizations should also have in place measures to evaluate and ensure compliance with, and effectiveness of, these policies.

- **Ensuring that necessary controls** are designed into AI agents and agentic AI systems. These could consist of privacy-by-design and privacy-by-default measures that facilitate the exercise of individuals' data protection rights (such as access and correction rights); requiring the agent or agentic AI system to maintain end-to-end traceability (such as records of tool calls, inputs and outputs); preferring agentic AI systems whose actions are designed to be repeatable or reliable; compartmentalizing agents' memory across users, or disabling in-memory storage where appropriate; establishing key decision points or actions (including, as necessary, interim points in a decision process before action is taken) which require an agent to seek human intervention or otherwise establishing limits to AI agents' or agentic AI systems' autonomy, and regularly evaluating the effectiveness of these measures; or, enabling users or operators to pause or stop execution, and where feasible and reasonably practicable, enabling actions to be reversed or systems and data to be restored.
- **Testing and real-time monitoring** of agentic AI systems. Testing ensures that agentic AI systems behave as expected before deployment, and could include functional testing, adversarial testing to identify vulnerabilities, and robust validation of safeguards such as data access controls, human intervention triggers, and reversibility mechanisms. Post-deployment, as agents interact with dynamic, real-time environments, including other agents, measures should be in place to monitor for and report on unexpected behaviors or failures.
- **Maintaining records** of relevant decisions, actions, tool calls, inputs, outputs and human interventions sufficient to support oversight, investigation and accountability. Logging and traceability measures should themselves comply with applicable privacy and data protection principles, including data minimization, purpose limitation, access controls, security safeguards and appropriate retention periods, recognizing that logs may contain prompts and sensitive personal data about users or third parties.
- **Maintaining governance over the deployment, configuration and use of agentic AI systems within the organization**. This could include identifying the parties respectively accountable for the design, use, and outcomes of the agentic AI system, maintaining an up-to-date inventory of approved systems and their purposes, models, tools, skills, data sources, connections and permissions; identifying unauthorised or materially altered deployments; establishing procedures to detect, investigate and respond to incidents; and updating relevant records following material changes.
- **Establishing proportionate mechanisms for challenge, human review and effective recourse.** Where an agentic AI system makes, recommends or carries out a decision or action that has a legal, material or similarly significant effect on an individual, organizations should

provide an accessible means to challenge the outcome and obtain meaningful review by a suitably qualified and authorized human decision-maker.

- **Implementing role-based AI literacy and awareness programs** for end users of agentic AI systems, as well as personnel involved in the governance, procurement, and deployment, senior management, and Data Protection Officers (DPOs), ensuring that each group has the knowledge and skills necessary to understand the capabilities, limitations, risks, governance requirements, and privacy implications associated with the deployment and use of AI agents, as well as their responsibilities when using agentic AI systems. This can include organization-specific knowledge, such as expectations or best practices with respect to structuring queries or defining tasks assigned to an agent, limits on the types of tasks for which an agent can be used, or requirements for the documentation that must be maintained when relying on an agent in a decision-making process. Such programs could also support the exercise of meaningful human oversight by enabling individuals responsible for reviewing, validating, challenging, or intervening in agentic AI outputs and decisions to do so effectively.
- **Providing meaningful notice, transparency and explanations to affected individuals.** Organizations should ensure that individuals are informed, in a clear, accessible and timely manner, when they are interacting with an AI agent, receiving AI-generated data, or being affected by a decision or action taken or materially supported by an agentic AI system. Data provided should be proportionate to the nature and potential impact of the processing and should enable individuals to understand the system's role, the personal data used, the principal factors or processes leading to the relevant outcome, and how responsibility for the system's actions can be traced.

The 48th Global Privacy Assembly therefore resolves to:

1. Call on organizations to consider the privacy risks set out in this resolution when designing or using agentic AI systems and encourage organizations developing agentic AI systems to research and implement innovative means of addressing privacy challenges.

2. Through the GPA Ethics and Data Protection in Artificial Intelligence Working Group continue to develop and share knowledge and best practices with respect to:
    a. the privacy considerations associated with agentic AI systems and how to mitigate them;
    b. instances in which organizations have deployed agentic AI systems to improve their data protection practices;
    c. the ways in which existing privacy principles can be effectively and predictably applied to agentic AI systems; and,
    d. the ways in which data protection authorities can facilitate the responsible development and use of agentic AI through a pro-innovation approach to regulation, including through the creation and operation of regulatory sandboxes.

3. Examine the potential for further work on this topic, including but not limited to developing guidance for organizations seeking to deploy agentic AI and awareness raising activities for the public (including guidance for individuals considering using agentic AI systems).

4. Continue to emphasize the importance of designing privacy protection into new and emerging technologies, particularly where they rely on the processing of personal data.

The United States Federal Trade Commission abstains from the adoption of this resolution.