Strasbourg, 2 September 2025 T-PD(2025)3

CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA

(CONVENTION 108)

Draft Guidelines on Privacy and Data Protection in the context of Large Language Models-based systems

www.coe.int/dataprotection

Outline

1. Purpose and Scope

- 1.1 Assist stakeholders in identifying, assessing, mitigating and monitoring (IAMM approach) privacy risks arising from LLM-based systems under the Articles and principles of Convention 108+.
- 1.2 Target audience:
    - States and Governments
    - LLM Providers (model developers and systems designers)
    - Deployers (system integrators and service providers)
    - Industry
    - Civil society organizations
    - Users (end-users)
    - Regulators / Supervisory Authorities

2. Key Definitions and Terminology

- 2.1 Definitions used throughout the document
    - 2.1.1 Difference between LLM Model and LLM-based System:
    - 2.1.2 Working definitions (Tech): AI foundation models, LLM, Small Language Models, AI agents, orchestration of LLMs, deployment, inference, training data, risks, etc.
- 2.2 Clarification of essential concepts in this context:
    - 2.2.1 Lifecycle of LLMs and its dynamic evolution (selected 5 fundamental steps)
    - 2.2.2 Risk management process
    - 2.2.3 Types of privacy risks
    - 2.2.4 Data protection by design and by default
    - 2.2.5 Mitigation strategies

3. Convention 108+ Principles Relevant to LLM-based systems

- 3.1 Outline the core data protection principles from Convention 108+ [Chapters 2 and 3]
- 3.2 Bipartite structure citing relevant Articles of the Convention and including in each sub-part (a) the interpretation and explanation of the principle, and (b) its contextualization.
    - 3.2.1 Lawfulness and fairness
    - 3.2.2 Purpose limitation
    - 3.2.3 Data minimization
    - 3.2.4 Data accuracy and quality
    - 3.2.5 Data subject's rights
    - 3.2.6 Transparency
    - 3.2.7 Data security
    - 3.2.8 Accountability, etc.

- Explain and contextualize how these principles relate specifically to Large Language Models and LLM-based systems and their lifecycle (e.g., during training, fine-tuning, or deployment.)

- Highlight specific tensions and challenges LLMs and LLM-based systems raise under these principles (e.g., repurposing data, hallucination, inferencing personal data, synthetic data risks).

4. Stakeholder-Specific Guidance

For each stakeholder group (providers, deployers, users, regulators):
1/ how each principle translates into concrete obligations or responsibilities for that group;
2/ how each actor is expected to have a methodology to identify, assess, mitigate and monitor privacy risks;
3/ with examples of privacy risks and their mitigations strategies;
Thus, providing (1) a mapping of Convention 108+ Principles to Actions, (2) the details of Risk Management Responsibilities and (3) examples of Mitigation Strategies and Best Practices

4.1. Mapping of Convention 108+ Principles to Actions
- Show how each principle translates into concrete obligations or responsibilities for that group

4.2. Risk Management Responsibilities: Identify, Asses, Mitigate and Monitor approach (IAMM)
- Indicate how each actor is expected to follow the IAMM stepped approach:

4.2.1 Identify privacy risks
4.2.2 Assess impact
4.2.3 Apply Mitigation strategies
4.2.4 Monitor and review effectiveness

4.3. Mitigation Strategies and Best Practices
- Example:
  - Providers: Data curation practices, differential privacy, robust evaluation metrics
  - Deployers: Use of consent mechanisms, prompt monitoring, and access controls
  - Users: Responsible prompt design, privacy-respecting use cases
  - Regulators: Oversight mechanisms, capacity-building, and audit criteria

5. General Recommendations and Implementation Considerations

5.1 Highlight governance and accountability mechanisms
5.2 Promote cross-functional collaboration (technical, legal, ethical teams)
5.3 Encourage human rights and fundamental rights impact assessments (complementarity of PIA and HUDERIA)
5.4 Point to interoperability with other regulatory frameworks (e.g., AI Act, GDPR, DSA)

6. Annexes

Annex 0: Justify why certain definitions were selected (e.g., alignment with AI Act, Convention 108+, OECD, ISO, etc.)

Annex I: Risk Management Process for LLMs

- Overview of risk identification, analysis, mitigation, monitoring
- Integration with data protection impact assessments (DPIAs)

Annex II: Evolving Lifecycle Phases of LLM Systems

- Detailed description of lifecycle stages relevant for LLMs

Annex III (optional): Case Studies or Examples

- Illustrative examples of privacy risks in real-world LLM deployments
- How different stakeholders addressed them
- Agentic AI

Annex IV (optional): Glossary

- List of key terms with brief definitions for easy reference

Introduction

Facing technological challenges to Privacy and Data Protection

The pace of technological development witnessed in the last years is gradually redefining the meaning of Private life and Data Protection. In the era of advanced AI foundation models and LLM-Based agentic systems these Guidelines aim at addressing the Privacy risks of these technologies and way they can interfere with individuals' rights.

As LLMs continue to shape the future of user-facing AI applications and gain traction across sectors such as recruitment, education, healthcare, and public administration, ensuring robust privacy protections becomes ever more challenging and the need to protect and uphold individuals' rights grows more urgent.

Audience and Structure

The development of comprehensive guidelines on the management of privacy and data protection risks under Convention 108+ provides Governments, Data controllers and Regulatory authorities, as well as Designers and Developers, Deployers and End-users (Part 1) with the necessary guidance and tools to identify, assess, and mitigate those risks. At the same time, it promotes compliance with privacy and data protection obligations (Part 4) within a consistent and clear terminological landscape (Part 2).

The Committee's role and its previous work in interpreting Convention 108+ in the context of emerging technologies, have been key to advancing regulatory clarity, strengthening international cooperation, and supporting research-informed policymaking (Part 3).

One of the core contributions of these Guidelines is to explain and contextualize how the principles of Convention 108 (cf. chapters 2 and 3) relate specifically to LLMs (and SLMs) and LLM-based systems throughout their lifecycle — during training and Model creation, Post-training adaptation and fine-tuning, System integration, Operational deployment and End-user interaction — as set out in Part 3.

Life-cycle approach: distinguishing LLM models and LLM-based systems

A life-cycle approach grounded in Convention 108+, the Framework Convention on AI and human rights standards guarantees that privacy and data protection principles are embedded throughout the development and use of AI systems, rather than introduced only at later stages as an afterthought.

Privacy and data protection risks of LLMs can emerge across different phases of an LLM model and system lifecycle phases (Part 2), ranging from model training and inference to integration and deployment within broader systems, having implications not only for data protection, but also for private life, dignity, and autonomy. The selected stages include:

(1) Model creation, where training data is collected, pre-processed and models are built with privacy-by-design considerations ;

(2) Post-training adaptation, where models are instructed, finetuned and transformed into assistance tools that are adapted to tasks ;

(3) System integration, in which LLMs are integrated into applications or services (often involving fine-tuning of pre-trained models) with appropriate safeguards;

(4) Operational deployment, referring to the live deployment of the LLM-based system with active monitoring and governance controls; and

(5) End-user interaction, covering how users interact with the LLM-based systems and how autonomous workflows or agentic functions are managed.

Ultimately, adopting a life-cycle approach will help ensure that as LLM technologies advance, they do so alongside robust privacy risk management (Part 4), achieving the necessary balance between innovation and compliance for the benefit of individuals and society at large.

Risk management: lifecycle-based methodology to assess and manage privacy risks associated with LLM-based systems

Building on the Expert Report "Privacy and Data Protection Risks in Large Language Models (LLMs)", the Guidelines draw on consolidated, science-based preliminary evidence and legal reasoning to support the Committee's leadership in advancing a future-proof normative framework and promoting a coherent and forward-looking approach to privacy governance in LLM-based systems.

Through their focus on Privacy Risk Management best practices, the Guidelines ensure that the principles of Convention 108+ are translated into practical standards and that emerging technologies are aligned with democratic values, fundamental rights, and Rule of Law within the global mission of the Council of Europe.

While Identification, assessment, mitigation and monitoring of the privacy and data protection risks associated with the use of Large Language Models (LLMs), need to evolve alongside technological advancements, these Guidelines also respond to the growing demand from organizations deploying LLM-based systems and agentic workflows for practical, interoperable frameworks that cover the AI lifecycle. To this end, these guidelines incorporate research and industry best practices, address common challenges, and highlight effective mitigation strategies relevant for key stakeholders including AI developers, deployers, researchers, policymakers, civil society organizations and regulators.

Governance Mechanisms

The present Guidelines will also promote a proactive, rights-based approach to innovation while safeguarding the principles of transparency, accountability, and human dignity at the heart of Convention 108+ and the Council of Europe's new Framework Convention on Artificial Intelligence. They provide transversal guidance on governance and accountability mechanisms that promote cross-functional collaboration between technical, legal, and ethical teams and encourage the use of human rights and fundamental rights impact assessments that are interoperable across regulatory frameworks (Part 5). In doing so, the Guidelines contribute to the joint standard-setting efforts of the relevant Council of Europe Committees in advancing an integrated approach to AI governance and data protection, complementing existing evaluation and governance methodologies such as Privacy Impact Assessments (PIAs) and HUDERIA, and offering a broader perspective centered on systemic impacts on human rights, democracy, and the rule of law.

Council of Europe: Emerging technologies and Private life

The Council of Europe is uniquely positioned to ensure that international AI Governance and Data Protection evolve in parallel, and that the protection of human dignity, privacy, and democratic oversight remains at the heart of the global technological progress thanks to its three complementary Conventions (Convention 108+, the Framework Convention on AI and the Convention on Cybercrime). These actions will support a robust, scalable framework that enables both human rights-centered innovation and accountability, while ensuring regulatory convergence and avoiding fragmentation.

As LLM-based systems continue to reshape the digital landscape, these guidelines offer specific guidance to all stakeholders on a shared methodology in privacy risk management (Part 4), offering an actionable interpretation of the principles enshrined in Convention 108+ (Part 3) and an international coordinated governance framework (Part 5), that ultimately can serve as a cornerstone for global alignment on privacy in the age of generative AI, also compliant with the Framework Convention on AI and capable of guiding responsible innovation across borders.

Open questions for discussion:

1. Regarding the title: “Guidelines on Privacy [Risks] and Data Protection in the context of LLM-based systems”

OR: “Guidelines on Addressing Privacy and Data Protection Risks in the context of LLM-based systems”

Any other suggestion for the title is also welcome.

2. In section 1.1.2. experts have proposed to include the term AI agents, and we would like to discuss how to frame this definition.

3. Should the justification the Guidelines will provide for the choice of definitions be part of the section 2. or be included in an annex?