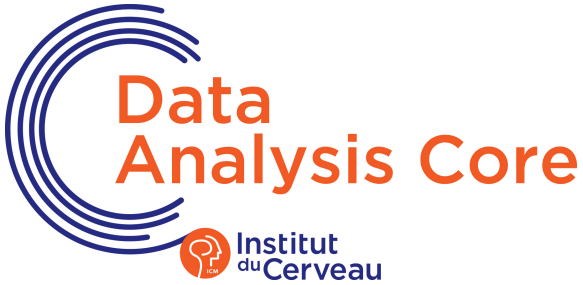

AI policy for responsible use of artificial intelligence

1 Introduction 1

2 Principles 2

3 Rules for specific areas 3
- 3.1 Sensitive and personal data . . . 3
- 3.2 Coding, data analysis, and reproducibility . . . 4
- 3.3 Scientific writing and publication . . . 5
- 3.4 Human and team impact . . . 5

4 If something goes wrong 6

5 Rationale and sources 6
- 5.1 Principles . . . 6
- 5.2 Scientific writing and publication . . . 6
- 5.3 Sensitive and personal data . . . 7
- 5.4 Coding, data analysis, and reproducibility . . . 9
- 5.5 Human and team impact . . . 11

6 Legal framework 13
- 6.1 EU AI Act . . . 13
- 6.2 Sensitive and personal data . . . 13
- 6.3 Health data reuse . . . 14
- 6.4 Institutional and funders AI policies . . . 14
- 6.5 Alignment with quality management . . . 15

A Disclosure statement templates 16

B Glossary 17

C Acronyms 18

D Bibliography 20

E About this document 26

1 Introduction

Artificial intelligence (AI), and generative AI in particular, is now part of everyday scientific work. It drafts and edits text, writes and explains code, summarizes literature, and assists with analysis. Used well, it can speed up discovery and lower barriers for the whole team. Used carelessly, it can leak confidential data, introduce fabricated results, erode skills, breach the integrity that research depends on. This policy sets out how the Data Analysis Core uses AI, so we benefit from it while protecting people, data, and the science.

The policy applies to everyone on the team: staff, students, interns, and contractors. It covers any AI tool used for work, at the office or when working from home, from general-purpose chatbots and coding assistants, to AI agents that can act on files and systems, and domain-specific models we might build or deploy. It covers four areas: scientific writing and publication; sensitive and personal data; coding, data analysis, and reproducibility; and the human and team impact. Each has its own set of rules in Section 3.

This policy sits underneath Paris Brain Institute's institutional and RSSI security policies and the wider legal framework, drawing them together into one summary for the team to work from (Section 5 and Section 6). It does not replace binding law or a stricter rule from Paris Brain Institute, its supervising bodies, funders, or journals. Where any of those impose a stricter requirement, that requirement governs.

This is a living document, revised to keep pace with the technology, its regulation, and the Data Analysis Core's own activities. It is maintained by the Operational Manager of the Data Analysis Core, who is the contact for questions about it. It is reviewed at least once a year, and sooner once the RSSI's AI documents now in preparation are approved, when a major legal or institutional change lands, or when the Data Analysis Core takes on a new method, tool, or activity this policy does not yet cover. Suggestions from any team member are welcome at any time.

2 Principles

The seven principles of responsible use of AI at the Data Analysis Core

1. **Accountability is never transferable.** You answer for anything you produce or approve with AI, including the parts it wrote. A tool cannot answer for accuracy, integrity, or harm.
2. **Protect privacy and data by default.** Keep personal, clinical, and confidential data out of public tools, and keep it safe. An AI agent that can act on files can corrupt or lose data, even where confidentiality is not at risk.
3. **Understand and verify what AI produces.** Use it only where the result can be checked, by you or by a colleague who can, and check it against primary sources before you rely on it. Using AI to learn or to try something unfamiliar is welcome, as long as anything you rely on has been checked.
4. **Be transparent.** Say when you have made significant use of AI in what you publish, in what you deliver to collaborators and clients, and in your code. Talk openly with each other about how you use these tools.
5. **Use AI for genuine need.** Reach for it because it fits the task, not to compensate for a lack in time, people, or skill. That lack has to be dealt with directly.
6. **Using AI is optional.** No one has to use it, and choosing not to carries no penalty. The exception is a role that explicitly involves using, building, or evaluating AI.
7. **Weigh equity and sustainability.** Do not judge people by whether or how much they use AI, and do not treat speed or output as the measure of good work. Keep use proportionate to real need, against its environmental and human cost.

These principles draw on established international AI governance frameworks (Section 5.1).

3 Rules for specific areas

3.1 Sensitive and personal data

- Never put identifiable or pseudonymised personal, clinical, or patient data, including special-category data such as health data, into a general-purpose or public AI tool (Section 5.3.1).
- Never use public services for confidential or sensitive work. For project-specific code, unpublished research questions, and meeting notes, use the DSI-managed AI platform. Identifiable or pseudonymised personal, clinical, or special-category data must never go into any general-purpose tool (Section 5.3.1).
- Treat AI agents that can act on files and systems as a data-safety risk. Work on a copy or in version control, keep backups, grant it the narrowest access it needs, and confirm you can review and undo its actions. Never automate a critical action, such as clinical data analysis, without human supervision (Section 5.3.3).
- Follow data minimisation, and prefer synthetic or anonymised data. Do not assume pseudonymised data is safe, because re-identification risk remains. Where realistic data is needed, prefer synthetic data, but never by feeding identifiable data into a public tool (Section 5.3.1). Anonymisation must meet CNIL and EDPB criteria (Section 6). When AI helps develop code that will later run on sensitive data, test it on invented or synthetic data on a local machine, so the real data is never shared with the tool (Section 5.3.1).
- Confirm that data reuse is permitted before using AI. Reuse of health data is governed by law. French clinical and biobank data falls under the *Code de la Santé Publique* (Section 6), and data from elsewhere falls under its own country's equivalent. The agreement for each dataset can add further, often stricter, conditions on AI use (Section 5.3.4).
- Be aware that third parties may run their own AI tools during interactions with us, such as an external notetaker summarising a meeting, which can expose confidential information beyond our control. If you use such a tool, say so and explain how the information collected is protected (Section 5.3.2).
- Never enter a password, API key, or other login credential into an AI tool's prompt. A credential typed into a public or general-purpose tool can be exposed the same way any other confidential content can, but unlike most data, its loss is immediately exploitable (Section 5.3.1).
- Never provide a meeting or conference recording, audio or video, to an AI tool without the consent of its organisers and participants (Section 5.3.1).
- When you need AI to analyse, translate, or improve text or code, share only the excerpt you need, not the whole document, to limit what a tool retains (Section 5.3.1).
- Use a dedicated translation tool, such as DeepL, rather than a general-purpose generative AI tool, for translation. A dedicated translation tool does not retain submitted text the way a generative AI tool can (Section 5.3.1).

- Do not try to bypass an AI tool's built-in safety or moderation limits, including through prompt injection, to make it do something it is designed to refuse (Section 5.3.1).

3.2 Coding, data analysis, and reproducibility

- Own the code the AI helped you write. AI-generated or AI-assisted code is read, understood, and tested by the person who submits it, exactly as if they wrote it by hand. Where you are not yet fluent in what the AI produced, ask a colleague who is to review it with you, and treat this as a chance to learn what the code does and why, and confirm it behaves as expected, including by checking its output against an established or hand-checked analysis. (Section 5.4.1)

- Check that any function, package, or flag it uses actually exists and behaves as claimed, including where it comes from, before you trust or install it (Section 5.4.7).

- Guard against data leakage and inflated results. When AI or machine learning is used in analysis, watch specifically for train-test contamination and other leakage, a pervasive cause of over-optimistic results that do not reproduce (Section 5.4.5).

- Document where AI significantly shaped an analysis pipeline, including the tool, its version, and the prompts or configuration that matter to the result. Keep this document alongside the code in version control so that another person, or the same person later, can reconstruct what was done. Where an AI agent carries out several steps on its own, document the tool calls and intermediate steps, not only the prompt that started it (Section 5.4.3).

- Label what AI generates. Data, results, and images produced by AI must be clearly marked as AI-generated and never presented as real observations, so no one mistakes them for measured data (Section 5.4.4).

- Do not use public tools for confidential code and data. The Data Analysis Core's general pipeline code is not confidential, but project-specific code, unpublished data, research questions, meeting notes, minutes of decision-making bodies, contractual documents, and collaboration-agreement material are, for confidentiality and strategic or competitive sensitivity rather than personal-data protection. The same safeguards apply as for personal data, and since no AI tool is currently validated for strategic, financial, or intellectual-property information, keep it out of AI altogether until a dedicated, validated environment exists (Section 3.1 & 5.3.1).

- Use only DSI-approved coding assistants. A coding assistant on a personal account is not permitted, and any AI extension installed in a code editor needs a DSI-identified professional licence (Section 5.4.1). Ask the DSI for an approved coding assistant and licence through an IT ticket before using one.

- Use verified, authorised training data, never unvetted material scraped from the internet. Document, version, validate, and archive data and code when building or fine-tuning a model (Section 5.4.5). Run it on DSI-managed infrastructure or computers. Check its integrity regularly for tampering, along with any scripts, binaries, or other files it depends on (Section 5.4.6).

- Use only AI models, container images, and packages registered and approved by the DSI. An open-source model or tool needs prior validation by DSI and Information Security before use. That review covers security, licensing, training data, and integration into our infrastructure (Section 5.4.2).

3.3 Scientific writing and publication

- AI is never an author. No AI tool is listed as an author or co-author on any team output, because authorship carries responsibility for accuracy, integrity, and originality, and for managing conflicts and copyright (Section 5.2.1).

- Do not put confidential peer-review material, including unpublished manuscripts, data, and grant applications, into any external or public AI tool, consistent with funder and publisher rules (Section 5.2.1).

- You are fully accountable for AI-assisted content, including every AI-assisted sentence, figure, citation, and dataset, and for checking for factual error, fabrication, and plagiarism (Section 5.2.1). Verify every reference against the primary source before you cite it, since AI-fabricated citations are common and passing one off as genuine is a research-integrity failure (Section 5.2.2).

- Disclose substantial AI use, whether the output goes to a journal or is shared with a client, colleague or management (Section 5.2.1 for journals). Basic grammar and spell-checking of human-written text does not need disclosure. Examples are provided in A.

- Be cautious with AI-generated images. Policies vary by publisher, so check the target journal's policy before using or including one. (Elsevier, 2026b; Springer Nature, 2026).

- Do not rely on AI to generate a project's core idea, whether a research question, an analysis approach, or a new tool's design. Its output only recombines existing material, so it cannot supply the originality the work needs (Section 5.5.1).

3.4 Human and team impact

- Protect skill formation, especially for junior staff, and require training before AI use. Room must remain to keep exercising underlying skills. Before using the DSI-managed AI platform, staff must complete the required training and access process. We encourage staff to take external learning opportunities and discuss AI use as a team, alongside the data-protection and integrity rules in this policy (Section 5.5.1).

- Maintain an honest, questioning culture around AI. It is acceptable, and expected, to challenge an AI-assisted result, because people tend to over-trust automated recommendations (Section 5.5.1). Mistakes and near-misses are shared openly, because these tools are new and go wrong in unfamiliar ways, and anyone can raise a problem without blame (Section 4).

- Be honest about wellbeing and workload. AI can remove tedium, but it can also increase the pace of work, so we introduce it with the team's agreement and check that it improves the work, not only its speed (Section 5.5.3).

- Do not use AI to cover for a lack of skill, staff, or time. A skills gap is addressed with training where that is possible (Section 5.5.1). A staffing or time gap is addressed through project management, prioritization, and recruitment. Where a gap affects the quality of the work, it is raised with the direction through the annual management review (Section 6.5).

- Do not let AI judge people. AI tools are biased so they stay out of any evaluation of colleagues, students, or applicants. Decisions about people, including hiring, promotion, and assessment, stay in human hands. Where AI helps with such decisions, such as recruitment or evaluation, disclose this to those affected and screen the content being assessed for hidden prompts before relying on its output (Section 5.5.2).

- Use AI proportionate, frugal, and limited to where it is genuinely needed. Prefer the smallest adequate model, favour a single model over a multi-agent chain, and use a lighter reasoning setting where that does the job (Section 5.5.4).

4 If something goes wrong

Mistakes and near-misses will happen, and reporting them quickly is what limits the harm. If personal, clinical, or confidential data is shared with a tool that should not have received it, or an AI agent has damaged or exposed data or code, act at once. Stop using the tool, and do not quietly delete the data or hide what happened, because that removes the record needed to assess and contain the exposure. Tell the Operational Manager promptly. Report it to the RSSI through the DSI ticket form or email rssi-cssi@icm-institute.org, and for anything involving personal or health data, notify the DPO at dpo@icm-institute.org, so the incident can be assessed and, where the law requires it, reported within the applicable deadline. Raising a problem early is responsible practice, and the institute's IT Charter (RSSI-REF-004, 2026, §9) protects good-faith reports from sanction, including from a person who caused the incident unintentionally, in keeping with the wider principle that institutions protect those who raise concerns in good faith (ALLEA, 2023, §3.2).

5 Rationale and sources

5.1 Principles

The seven principles in Section 2 draw on the international AI governance frameworks set out by the OECD AI Principles (OECD, 2024) and the UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021), alongside the ERA Living Guidelines (European Commission & European Research Area Forum, 2026) and the ALLEA code of conduct (ALLEA, 2023).

5.2 Scientific writing and publication

5.2.1 Publisher and funder rules

Research funders and publishing houses impose their own rules, and these are consistent with each other. The European Research Council (ERC) prohibits AI in proposal evaluation (European Research Council, 2026), Wellcome and the broader ERA framework require responsible use and disclosure (European Commission & European Research Area Forum, 2026; Wellcome Trust, 2023), and CoARA sets the

direction for how AI is used to evaluate people and outputs (Coalition for Advancing Research Assessment, 2025). National Institutes of Health (NIH) prohibits generative AI in grant peer review and limits AI-developed applications (National Institutes of Health, 2023, 2025c). Before submission or review therefore check each funder's rules (COMETH, 2024). Across the ICMJE (International Committee of Medical Journal Editors, 2024), COPE (Committee on Publication Ethics, 2023), WAME (World Association of Medical Editors, 2023), Springer Nature (Springer Nature, 2026), Science (American Association for the Advancement of Science, 2026), Elsevier and Cell Press (Cell Press, 2026; Elsevier, 2026b), and PLOS (Public Library of Science, 2026), publishers agree that AI cannot be an author, that a human author is accountable for AI-assisted content, that material use must be disclosed, and that confidential manuscripts must not be uploaded to AI tools during review (Elsevier, 2026a; Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique, 2026). Because Cell Press and other neuroscience journals are part of Elsevier, their requirements apply to most journals our work is published in. Inria, RSSI and COMETH reach the same position on authorship and accountability (Arias et al., 2025; COMETH, 2024; Paris Brain Institute, RSSI [RSSI-DOC-003], under review).

5.2.2 Citation integrity

AI regularly invents references that look real but do not exist (Naddaf & Quill, 2026). The ECoC treats concealed AI use as research misconduct (ALLEA, 2023; Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique, 2026, §3.1), and passing off an AI-fabricated citation as genuine, rather than checking it against the primary source, falls under that same standard.

5.3 Sensitive and personal data

5.3.1 Data in public AI tools

The IT Charter (RSSI-REF-004, 2026, §5) prohibits putting personal or special-category data, including pseudonymous data, into a public or general-purpose AI. Content entered there can be reused to train the model and become accessible to other users, and the leak cannot be undone. The RSSI's staff security guide (2024, §2.5) extends this prohibition to every stage a system built on AI touches, not only the prompt: training data, test data, any data added after deployment, and any vector database it draws on. Paris Brain Institute's Data Policy already requires that data be pseudonymised before research use with the institute's in-house Identity Manager tool (Corvol et al., 2024). Whether pseudonymised data may be used in an AI task depends on the same legal basis and impact assessment requirements as any other AI use of personal data (Section 6). Inserm's and Stanford's AI governance guidance set out a similar tiered data-classification approach, under which high-risk data is barred from tools lacking an institutional agreement (Inserm, 2025; Stanford University, 2025). Inserm goes further: it prohibits non-sovereign external services outright. COMETH's opinion names further examples of data that must stay out of an external tool, including minutes of decision-making bodies and contractual documents (COMETH, 2024). It also warns against generating synthetic data by feeding real identifiable data into a public tool to have it invent something similar: the real data has already been exposed to the tool by that point, regardless of what the output looks like (COMETH, 2024). Before adopting a new tool, COMETH suggests to look at Ithaka S+R generative AI product tracker, which catalogues AI tools of interest to higher education (COMETH, 2024). RSSI-DOC-003 adds further practical safeguards for public and general-purpose tools: never enter a password or other credential into a prompt, and do not attempt to bypass a tool's built-in safety or moderation limits. Where the sensitivity of the work warrants it, turning off chat history and limiting a tool's file and contact permissions further reduces what it retains beyond

the session (RSSI-DOC-003, under review). Sorbonne Université's Faculté de Santé recommendations name further examples: never provide a meeting or conference recording without the organisers' and participants' consent, share only short excerpts rather than whole documents when analysing, translating, or improving text or code, and use a dedicated translation tool rather than a general-purpose one, since it will not retain submitted text the same way (Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique, 2026). CLUSIF's security model proposes a similar four-tier classification tying data sensitivity to where a tool runs, from the institute's own infrastructure through EU-hosted to non-EU-hosted services (Club de la Sécurité de l'Information Français [CLUSIF], 2025).

5.3.2 Third parties' own AI use

Third parties we work with, including collaborators and clients, may run their own AI tools during interactions with us. Note-taking assistants that summarise a meeting or draft a follow-up carry the same confidentiality, data-protection, and intellectual-property risks this policy sets for our own use, but outside our control. The ERA Living Guidelines ask that we stay mindful of this and inform third parties of the risk, for example by sharing this policy or the relevant guidance, and, where we are the one using such a tool, that we tell them and explain how the information collected is protected (European Commission & European Research Area Forum, 2026, p. 7).

5.3.3 Containing AI agents

An AI agent that can read, write, run, move, or delete files or call other tools can cause damage beyond what the content it is shown would suggest. Containing what it can do matters as much as vetting the information it has access to. This risk covers integrity as well as confidentiality: an agent that acts on files can corrupt or lose data even where confidentiality is never at risk. GDPR requires that personal data be kept accurate and secure against this kind of loss or alteration (European Parliament and Council of the European Union [European Union], 2016, Art. 5(1)(f) and 32). The institute's IT Charter (RSSI-REF-004, 2026, §7) states this directly: critical actions on the infrastructure, such as clinical data analysis, network reconfiguration, or creating a privileged account, must not be automated with an AI tool or script that runs without human supervision, because an algorithmic error could compromise the integrity of research results or of the information system itself. Recent French policy guidance on agentic AI recommends going further: running an agent in a sandboxed environment, clearly separating what a user can instruct it to do from what the system itself allows, keeping any data it retains to a strict minimum and duration, and regularly testing that an intervention actually stops the agent before it is trusted with more autonomy (Inria, 2026). RSSI-DOC-003 makes the same point about privilege generally. An AI system should run with the minimum privileges it needs, such as in a restricted container or virtual machine, rather than with administrator rights (RSSI-DOC-003, under review).

5.3.4 Data reuse conditions

Data providers often restrict AI use. UK Biobank bars incorporating participant-level data into a public generative AI model, including indirect exposure such as posting it to a public code repository. It treats a model that can reproduce or regenerate participant-level data as equivalent to giving direct access to it. Therefore, such a model cannot be shared even where it was trained privately (UK Biobank, 2026). The NIH has similarly paused sharing generative AI models, and their parameters, trained on its

controlled-access genomic data, beyond a project's closeout, pending further guidance on the privacy risk this creates (National Institutes of Health, 2025a, 2025b).

5.4 Coding, data analysis, and reproducibility

5.4.1 DSI-licensed coding assistants

The RSSI's software development security policy (RSSI-REF-010, 2024, §T, T.01–T.02 and T.08) requires that coding assistants run only through tools and licenses the DSI provides, bars access through personal accounts for the institute's software development, and requires that any AI extension in a code editor carry a DSI-identified professional license. Inria reached a similar conclusion, and pending a sovereign option, recommends that institutions facilitate licensed access to generative AI tools, rather than leaving staff to rely on personal, uncontrolled accounts (Arias et al., 2025). On coding specifically, Inria goes further. Citing the many uncertainties still surrounding generative AI, it recommends restricting its use in programming to non-critical code (Arias et al., 2025). This is close to a distinction the RSSI already draws, from a different angle: the IT Charter (RSSI-REF-004, 2026) bars automating a critical action, such as clinical data analysis, with an AI tool or script that runs without human supervision (Section 5.3.3). CLUSIF's security model adds a specific safeguard for AI-generated code: it must never run or be committed automatically, and needs a scanning tool and human review first, which is what the ownership rule in Section 3.2 already requires (CLUSIF, 2025). Inria also notes that code hosted on a collaborative platform, such as GitHub, can end up training other organisations' models without regard for that code's own licence terms, and asks staff to bear this in mind when choosing where to host development work (Arias et al., 2025). We do not additionally sort code itself into critical and non-critical categories, since that boundary is hard to draw cleanly and the RSSI's supervision requirement, together with the licensing requirement above, already covers the underlying risk. Finally, regarding authorship and accountability, Inria and COMETH's opinion hold a human accountable for AI-assisted output generally, code included, not only manuscript text (Arias et al., 2025; COMETH, 2024).

5.4.2 Vetting third-party models

A downloaded model or component carries a risk distinct from AI-generated code. A malicious or compromised model can be planted in a public repository. Some formats, such as pickle, can execute arbitrary code the moment the model is loaded. RSSI-DOC-003 requires DSI and RSSI to review any open-source model or tool before use, covering its security, licensing, training data, and integration into our infrastructure. It restricts staff to models, images, and packages already registered and approved by DSI (RSSI-DOC-003, under review). To have a new tool or model assessed, raise it with the DSI's application team through an IT ticket, which is referred to the RSSI where its use needs a security review. If you are unsure whether a tool is already validated, ask the RSSI.

5.4.3 Recording AI's role in analysis and code

The ECoC's transparency standard treats concealed AI use as misconduct (ALLEA, 2023, §3.1), and this applies to computational work as much as to writing. It requires documenting where AI significantly shaped code or an analysis pipeline, so the work can be reconstructed later. French policy guidance on agentic AI recommends that records are also maintained of tool calls and reasoning steps, and to versioning the decoding parameters and random seeds involved, not only the prompt that started it (Inria, 2026).

5.4.4 Labelling AI-generated content

The EU AI Act requires that a person be told when they are dealing with AI, and that synthetic content be labelled (European Union, 2024, Art. 50). Inserm's recommendations set the same expectation for data, results, and images produced by AI. These must be marked as AI-generated and never presented as real observations (Inserm, 2026).

5.4.5 Reproducibility and computational integrity

AI coding assistants have been shown to often produce code that passes obvious tests, but fail on less common inputs and edge cases that can quietly change a result later (Liu et al., 2023). The coding rules in Section 3.2 address this.

Where the team builds or fine-tunes a model rather than using existing tools, Inserm's recommendations set out what reproducibility requires: documenting the training data and any known biases, recording the model version, validating that a model performs as intended before it is relied on, and archiving models so a result can be checked and repeated later (Inserm, 2026). Where that training data includes personal data, the CNIL's GDPR-specific development framework applies as well (Section 6). RSSI-DOC-003 requires that training data be verified and authorised before use, particularly for models handling regulated data such as health information (RSSI-DOC-003, under review). Machine-learning analyses carry a related risk even without training a new model: leakage between training and test data can inflate apparent accuracy and produce results that do not reproduce. Results must therefore be validated on properly held-out data before they are trusted (Kapoor & Narayanan, 2023). The IT Charter (RSSI-REF-004, 2026, §5) also requires that an AI model's integrity, and that of its training data, be checked regularly, so that tampering is caught before it silently changes a result, consistent with CLUSIF's requirement to hash or sign model files for the same purpose (CLUSIF, 2025).

5.4.6 Model infrastructure, integrity, and openness

The RSSI's staff security guide (2024, §2.5) keeps generative AI, including any model the team trains or fine-tunes, off personal computers. RSSI-DOC-003 sets out where it can run instead: DSI-managed infrastructure or computers (RSSI-DOC-003, under review). Beyond this security requirement, several sources converge on a broader preference for open and locally run tools where there is a reasonable choice. The ERA Living Guidelines recommend running locally hosted or self-governed AI (European Commission & European Research Area Forum, 2026), Inserm's guidance treats openness and preferring the smallest adequate model as part of responsible AI use (Inserm, 2026), and COMETH's opinion reaches the same preference (COMETH, 2024). Recent French policy guidance on agentic AI makes a related sovereignty point at the model level, alongside its efficiency benefit (Inria, 2026), and Inria recommends, pending a sovereign option, that institutions facilitate licensed access for their teams rather than leaving staff to rely on personal, uncontrolled accounts (Arias et al., 2025). This preference fits the team's own open-science and Findable, Accessible, Interoperable & Reusable (FAIR) practice, echoed in 2025 by the Council for National Open Science Coordination, which called on policymakers to align AI development with open-science principles and to use those principles to rein in untrustworthy AI practices (Council for National Open Science Coordination [CoNOSC], 2025). Where code or a model is published, the Data Policy already requires it to carry a distribution and usage licence and recommends submitting the published version to Software Heritage so it stays citable and permanently available (Corvol et al., 2024). This preference is conditional. An open-source model or tool still needs DSI and Information Security validation before use (Section 3.2) (RSSI-DOC-003, under review).

A further sovereignty risk applies specifically to tools hosted outside the EU. The US CLOUD Act of 2018 lets US law enforcement compel a US-based provider to disclose data it holds, even data stored on servers outside the United States, as long as the provider retains possession, custody, or control of it (U.S. Department of Justice, 2018). Data submitted to a US-headquartered generative AI provider can therefore reach US authorities regardless of where that provider's infrastructure sits. Such a disclosure may not meet the GDPR's own conditions for transferring personal data to a non-EU authority (European Union, 2016, Art. 48). The Sorbonne Université Faculté de Santé's recommendations raise the same risk (Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique, 2026).

5.4.7 Hallucinated functions and packages

AI regularly invents plausible-sounding function names, package interfaces, and command-line flags that do not actually exist, so code that looks right can fail as soon as it runs (Liu et al., 2023). This extends to the packages an AI suggests installing: a code-generating AI can invent a plausible-sounding package name that does not exist, which an attacker can then register and fill with malicious code (Spracklen et al., 2025). The RSSI's software development security policy (RSSI-REF-010, 2024, §T, T.07) is consistent with this: any suggested library must come from a trusted source before it is added.

5.5 Human and team impact

5.5.1 Skill formation and over-reliance

Over-reliance on AI correlates with reduced critical thinking and cognitive effort, especially for people still building their expertise (Gerlich, 2025; Lee et al., 2025; Swiss Young Academy, 2025). In addition, people tend to over-trust automated recommendations and accept a result without challenging it (Romeo & Conti, 2025). COMETH's opinion cautions against using AI opportunistically or as a substitute for lacking resources or expertise (COMETH, 2024).

The effect of AI on skills and wellbeing depends on how an organisation introduces it. The OECD finds that consulting staff and investing in training are most associated with fair outcomes (OECD, 2023). ERA Guidelines echos this in a call on research organisations to provide training (European Commission & European Research Area Forum, 2026). The Swiss Young Academy's recommends to provide training and the right tools to work with AI, and set incentives that support considered use, so that AI does not come at the cost of critical thinking, originality, and scientific rigour (Swiss Young Academy, 2025).

Inserm provides a course list (Inserm, 2026), drawn mainly from mentor.gouv.fr and FUN-MOOC, and tutorials from Creatis's Myriad team (creatis-myriad.github.io), covering autoencoders, graph neural networks, and diffusion models in depth. At Paris Brain Institute, staff must attend the AI Masterclass, review COMETH's opinion, and complete an access and commitment form through the ticketing tool (COMETH, 2024) before using the DSI-managed AI platform.

The Sorbonne Université Faculté de Santé's recommendations reach the same conclusion. Generative AI output only recombines existing material, so relying on it to generate a project's core idea would compromise the work's originality (Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique, 2026).

5.5.2 Evaluating people

Tools that detect AI-generated text are unreliable and discriminate against non-native English writers, so they should not be used to judge people (Liang et al., 2023). AI tools used to evaluate people carry other documented biases, including gender and race and age (Chen, 2023). Inserm and Inria recommend against substantial use of generative AI in evaluating people, such as peer-reviewing articles or assessing a researcher's dossier for promotion and to keep decisions about people in human hands (Arias et al., 2025; Inserm, 2025). Inserm also expects that any AI use in decisions about people, including recruitment or evaluation, be made known to those it affects (Inserm, 2026). Content submitted for evaluation, such as a grant application or manuscript, can also carry a hidden prompt, an instruction invisible to a human reader but readable by an AI tool, intended to steer an automated assessment. Where AI assists with evaluation at all, this risk must be screened for rather than trusting submitted content at face value (European Commission & European Research Area Forum, 2026, p. 12).

5.5.3 Work pace and intensity

AI adoption is linked to a faster pace of work. In the OECD's surveys of workers in finance and manufacturing, most AI users said AI had increased the pace at which they perform their tasks, which the authors note can indicate increased work intensity rather than only productivity (Lane et al., 2023). Surveys run since generative AI show the same. Most employees using AI report that the tools have added to their workload even as executives expect them to raise productivity (Upwork Research Institute, 2024). A large Danish study found that chatbots create new job tasks, such as reviewing AI output, and that most users put the time saved into other tasks rather than shorter hours (Humlum & Vestergaard, 2025). The OECD and the Danish study both found the best outcomes where employers consulted and trained their workers, the same protective factors already established for skill formation (Section 5.5.1).

5.5.4 Environmental cost

The computation behind large AI models has a rising energy and water footprint. Data-centre electricity use reached an estimated 415 TWh in 2024, around 1.5% of world consumption, and is projected to more than double to around 945 TWh by 2030, with AI the most important driver of that growth (International Energy Agency, 2025). Audited numbers now exist at the scale of a single model. Mistral AI's lifecycle analysis of its Large 2 model measured 20.4 $ktCO_2e$ and 281 000 m^3 of water for training and the first 18 months of use, and around 1.14 gCO_2e and 45 mL of water for a typical 400-token response (Mistral AI, 2025). The same analysis found that impacts scale with model size, a model ten times larger generating impacts an order of magnitude larger for the same output (Mistral AI, 2025). Inserm's recommendations consider efficiency, including using the smallest adequate model, as part of responsible AI use (Inserm, 2026). In addition, favouring a single model over a multi-agent chain, and a lighter reasoning setting where either does the job, both reduce energy cost (Inria, 2026). The human cost matters as well: work such as labelling harmful content to make a model safe to use is often outsourced and poorly paid (Arias et al., 2025; Perrigo, 2023). Paris Brain Institute has not set a formal carbon or energy limit for AI use, so proportionality here is a matter of judgement rather than a fixed threshold.

6 Legal framework

6.1 EU AI Act

The EU AI Act entered into force on 1 August 2024 and applies in stages (European Union, 2024, Art. 113). The Digital Omnibus, an amending regulation in force since July 2026, postponed the high-risk obligations, which now apply from 2 December 2027 for standalone high-risk systems and from 2 August 2028 for AI embedded in regulated products such as medical devices (Freshfields, 2026). It classifies AI systems into risk tiers, with obligations that rise with the risk. It sets transparency duties for generative and interactive systems, which apply from 2 August 2026, so a person must be told when they are dealing with AI and synthetic content must be labelled (European Union, 2024, Art. 50). Organisations that deploy AI must also support their staff's AI literacy, a duty in place since February 2025 (European Union, 2024, Art. 4), which the training requirement in Section 3.4 serves. AI that functions as a medical device, or that is used for diagnosis or clinical decisions, is likely to fall in the high-risk tier (European Union, 2024, Art. 6 and Annex III), which carries obligations on data governance, documentation, logging, human oversight, and robustness (European Union, 2024, Arts. 8–15). The Act exempts AI developed and used purely for scientific research (European Union, 2024, Art. 2(6)), but that exemption falls away once a system is deployed on patients or placed on the market. As AI agents that act with less step-by-step supervision come into wider use, a 2026 French policy assessment recommends a targeted review of how the Act's prohibited-practices, oversight, transparency, and systemic-risk provisions (European Union, 2024, Arts. 5, 9, 14, 15, 50, and 55) apply to multi-agent systems specifically (Inria, 2026).

6.2 Sensitive and personal data

Sensitive and personal data carry the greatest legal and ethical exposure for Data Analysis Core, and the data rules in Section 3.1 are strict. The GDPR is the baseline for any processing of personal data. Health data is a special category and needs a specific legal basis, such as explicit consent or a scientific-research basis with appropriate safeguards (European Union, 2016, Arts. 9 and 89). The EDPB has clarified, in its Opinion 28/2024 (European Data Protection Board, 2024), that a model trained on personal data cannot be assumed to be anonymous, that any reliance on legitimate interest must be documented through a balancing test, and that health data needs a special-category basis regardless. The CNIL issued detailed recommendations through 2024 and 2025 on applying the GDPR to AI (Commission Nationale de l'Informatique et des Libertés, 2025). In January 2026 it published a more specific, step-by-step framework for developing or training an AI system on personal data, covering purpose definition, choice of legal basis, data minimisation, retention periods, informing individuals, enabling their rights, security, and how to assess whether a trained model itself counts as personal data, together with a practical compliance checklist (Commission Nationale de l'Informatique et des Libertés, 2026). Any processing of personal data with AI needs a documented legal basis, and high-risk processing also requires a Data Protection Impact Assessment (DPIA) (European Union, 2016, Art. 35), called a PIA in the Paris Brain Institute Data Policy (Corvol et al., 2024). The RSSI's procedure for AI impact assessment and risk management, in preparation, will set out how that assessment is run for AI at Paris Brain Institute (Paris Brain Institute, RSSI [RSSI-DOC-00X], in preparation). The DPO, CDO, DSI and its security team (RSSI), can advise before such processing begins.

6.3 Health data reuse

The EHDS regulations entered into force on 26 March 2025. Most secondary-use provisions apply from March 2029, four years after entry into force, with the remaining data categories, including genomic data, following from March 2031 (European Commission, 2026). It is the emerging lawful route for accessing population-scale health data for research and AI training. Secondary use will require a data permit (European Union, 2025, Art. 68), will happen inside secure processing environments with no download of identifiable data (European Union, 2025, Art. 73), and prohibits re-identification and any use detrimental to the people in the data (European Union, 2025, Art. 54). Patients hold opt-out rights (European Union, 2025, Art. 71). The CSP governs French clinical and biobank activity, and will continue to do so. Where AI informs a clinical decision or acts as a medical device, a qualified human must stay in control, under those high-risk duties and the human-oversight requirement of the *Code de la Santé Publique* (Article L4001-3).

The Council of Europe Framework Convention on AI (Council of Europe, 2024) will add human-rights, privacy, and transparency duties as it is ratified. The European Union ratified it as an organisation in May 2026, but France does not yet appear among the individual ratifying states.

6.4 Institutional and funders AI policies

This policy sits underneath Paris Brain Institute's 2020 Ethics and Professional Conduct Charter and Data Policy (Corvol et al., 2024; Paris Brain Institute, 2020), the RSSI's internal security policies, and the framework of Paris Brain Institute's supervising bodies, Inserm, CNRS, and SU, aligning with the recommendations of Paris Brain Institute's ethics committee (COMETH, 2024). Three RSSI security policies are already in force: the staff security guide (2024), the general IT Charter (RSSI-REF-004, 2026), and the software development security policy (RSSI-REF-010, 2024). Four more are in preparation and validation and will govern once approved: the AI security good-practice guide (RSSI-DOC-003), the procedure for AI impact assessment and risk management (RSSI-DOC-00X), the IT Charter for the use of AI (RSSI-REF-00X), and the security policy for the use of AI (RSSI-REF-015). Where any RSSI document and this policy differ, the RSSI document prevails, and this policy stands as its practical companion for the Data Analysis Core.

Guidance from bodies such as Inserm, Inria, and CLUSIF, cited throughout Section 5, is advisory by comparison. The ANR states no position in its 2026 call guide or its charter on deontology and scientific integrity (Agence Nationale de la Recherche, 2024, 2026). The CNRS's Scientific Council called in 2023 for a working group to study the issue, rather than setting rules itself (CNRS Conseil scientifique, 2023). The CNRS has since given its staff a sanctioned tool: in December 2025 it deployed Emmy, a secure conversational AI built on Mistral AI's open models, hosted in European datacentres, covered by the GDPR and the EU AI Act, and configured so that staff data is not reused to train the models (CNRS, 2026). It offers Emmy as a controlled alternative to public tools and bars services such as ChatGPT and Gemini at work. SU's Faculté de Santé ethics and scientific-integrity commission published its own recommendations in April 2026, covering much the same ground as this policy, down to practical limits such as sharing only short excerpts with a tool rather than whole documents and using a dedicated translation service (Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique, 2026). The MESR has not published an official position either: a charter circulating on its website is a university-led draft still under construction, without ministry authority. Paris Brain Institute currently has no institute-wide AI policy yet, though its Data Policy already anticipates that AI-specific codes of conduct will need to be respected once they exist, without itself setting one (Corvol et al.,

2024). This document is the Data Analysis Core's own attempt to set one for its activities, and will be revisited if a broader institutional or ministerial policy follows. CLUSIF's model, built on NIST's AI Risk Management Framework and the MITRE ATLAS threat base, is one example of how that broader institutional policy could eventually take shape (CLUSIF, 2025).

6.5 Alignment with quality management

This policy runs through the Data Analysis Core's existing ISO 9001 (2015) quality-management system, giving traceability across the training, development, and expertise-building it requires. Mistakes and incidents are reported and tracked as Section 4 sets out, feeding the same annual management review as the rest of the Data Analysis Core's quality and risk work. A dedicated AI management-system standard, ISO/IEC 42001 (ISO/IEC 42001, 2023), already exists and could be adopted in the future.

A Disclosure statement templates

These are starting points to adapt. The exact statement and location depends on the report or journal, so check its policy before submission. ICMJE, COPE, PLOS, and Springer Nature accept a statement in the Methods describing the tool, its version, the date, and how it was used. Elsevier and Cell Press want the declaration immediately above the references. Science wants it in the cover letter, Methods, and Acknowledgements.

General manuscript preparation

During the preparation of this work the author(s) used [tool and version] on [date(s)] in order to [for example, improve the language and readability of the manuscript]. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Language editing only

The author(s) used [tool and version] for language editing, that is grammar and phrasing, of author-written text. No content was generated by the tool, and the author(s) take full responsibility for the final text.

Code and data analysis

Part of the analysis code was written with the assistance of [tool and version] on [date(s)]. All AI-assisted code was read, tested, and verified by the author(s), including against [held-out data / an independent hand-checked analysis], and the author(s) take full responsibility for its correctness.

Elsevier and Cell Press

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used [tool and version] in order to [reason]. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

No AI use

The author(s) declare that no generative AI tools were used in the preparation of this manuscript.

B Glossary

AI agent An AI system that does not only produce text but can also take actions on its own, such as reading, writing, running, moving, or deleting files, calling other tools, or browsing, in order to carry out a task with limited step-by-step human control. Its ability to act makes it more capable, and also more able to cause unintended damage.. 1, 3, 8

anonymised data Information that does not relate to an identified or identifiable person, or personal data rendered anonymous such that the person is no longer identifiable. Anonymised data falls outside the GDPR's scope (European Union, 2016, Recital 26).. 3

data minimisation A GDPR principle requiring that personal data collected or used be adequate, relevant, and limited to what is necessary for the purpose (European Union, 2016, Art. 5(1)(c)).. 3

generative AI AI that creates new content, such as text, code, or images, in response to a prompt. The chatbots and coding assistants most people have used are generative AI.. 1

machine learning A family of methods in which a model learns patterns from data rather than following rules written by a person. Most of today's AI, including large language models, is built with machine learning.. 4

pickle A Python format for saving objects to a file. It can execute arbitrary code when the file is loaded, so a pickle file from an untrusted source is a security risk. See Python's own documentation.. 9

pseudonymised data Personal data processed so it can no longer be attributed to a specific person without additional information, which is kept separately and protected. Because that additional information can in principle restore the link, pseudonymised data remains personal data under the GDPR (European Union, 2016, Art. 4(5)).. 3

special-category data A GDPR term for personal data needing extra legal protection beyond ordinary personal data, including health data, genetic and biometric data, and data revealing racial or ethnic origin, religion, or sexual orientation (European Union, 2016, Art. 9(1)).. 3

vector database A database that stores content as numerical vectors capturing its meaning, so a search finds the stored items closest in meaning to a query rather than an exact keyword match. It is the standard component behind an AI system that searches internal documents or data before answering, and it can retain information about the original content even though it does not store that content directly.. 7

C Acronyms

ALLEA All European Academies. 6

ANR *Agence Nationale de la Recherche*, France's national research funding agency. 14

CDO Chief Data Officer. 13

CLUSIF *Club de la Sécurité de l'Information Français*. 8, 9, 14, 15

CNIL *Commission Nationale de l'Informatique et des Libertés*. 3, 10, 13

CNRS *Centre national de la recherche scientifique*. 14

CoARA Coalition for Advancing Research Assessment. 6

COMETH *Comité d'éthique*. 7, 9–11

COPE Committee on Publication Ethics. 7, 16

CSP *Code de la Santé Publique*. 14

DPIA Data Protection Impact Assessment. 13

DPO Data Protection Officer. 6, 13

DSI *Direction des Systèmes d'Information*. 3–5, 9–11, 13

ECoC European Code of Conduct for Research Integrity. 7, 9

EDPB European Data Protection Board. 3, 13

EHDS European Health Data Space. 14

ERA European Research Area. 6, 11

ERC European Research Council. 6

FAIR Findable, Accessible, Interoperable & Reusable. 10

GDPR General Data Protection Regulation. 8, 10, 13, 14

ICMJE International Committee of Medical Journal Editors. 7, 16

Inria *Institut national de recherche en sciences et technologies du numérique*. 7, 9, 12, 14

Inserm *Institut national de la santé et de la recherche médicale*. 7, 10–12, 14

MESR *Ministère de l'Enseignement Supérieur et de la Recherche*, France's Ministry of Higher Education and Research. 14

NIH National Institutes of Health. 7

OECD Organisation for Economic Co-operation and Development. 6, 11, 12

PLOS Public Library of Science. 7, 16

RSSI *Responsable de la Sécurité des Systèmes d'Information*, ICM's IT security officer and team. 1, 6, 7, 9–11, 13, 14

SU Sorbonne Universite. 14

UNESCO United Nations Educational, Scientific and Cultural Organization. 6

WAME World Association of Medical Editors. 7

D Bibliography

Agence Nationale de la Recherche. (2024). *Charte de déontologie et d'intégrité scientifique*. Retrieved July 13, 2026, from https://anr.fr/fileadmin/documents/2025/Charte-deontologie-et-integrite-scientifique-ANR-2024.pdf

Agence Nationale de la Recherche. (2026). *Appel à projets générique 2026: Guide de l'aapg*. Retrieved July 13, 2026, from https://anr.fr/fileadmin/aap/2026/anr-aapg-2026-guide-v1.0.pdf

ALLEA. (2023). *The european code of conduct for research integrity, revised edition 2023*. ALLEA. https://doi.org/10.26356/ECOC

American Association for the Advancement of Science. (2026). *Science journals: Editorial policies (artificial intelligence)*. Retrieved July 12, 2026, from https://www.science.org/content/page/science-journals-editorial-policies

Arias, S., Bergmann, M., Campillo, F., Enard, M.-A., Fabre, C., Garcia, F., Guedj, B., Jeannot, E., Neglia, G., Peurichard, D., Racoceanu, D., Sagot, B., & Tworkowski, G. (2025). *Réflexions sur l'usage de l'IA générative pour les métiers de la recherche* (Approved by Inria's Commission d'Évaluation, 9 July 2025). Inria, Commission d'Évaluation. Retrieved July 13, 2026, from https://inria.hal.science/hal-05187992v1

Cell Press. (2026). *Cell press journal policies: Ai and ai-assisted technologies*. Retrieved July 12, 2026, from https://www.cell.com/structure/information-for-authors/journal-policies

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, *10*, 567. https://doi.org/10.1057/s41599-023-02079-x

Club de la Sécurité de l'Information Français. (2025). *Modèle de politique de sécurité des systèmes d'information (pssi) pour l'intelligence artificielle* [February 2025]. Retrieved July 24, 2026, from https://clusif.fr/wp-content/uploads/2025/09/20250203-Clusif-PSSI-IA.pdf

CNRS. (2026). *Le cnrs lance avec mistral ai un outil d'ia générative sécurisé mis à disposition de ses personnels* [Emmy, deployed 17 December 2025, announced 3 February 2026]. Retrieved July 27, 2026, from https://www.cnrs.fr/fr/actualite/le-cnrs-lance-avec-mistral-ai-un-outil-dia-generative-securise-mis-disposition-de-ses

CNRS Conseil scientifique. (2023). *Usages et développements de l'intelligence artificielle générative et démarche scientifique*. Retrieved July 13, 2026, from https://www.cnrs.fr/comitenational/cs/recommandations/24-25_avril_2023/CS_usages_et_developpements_IA_generatives_demarche_scientifique.pdf

Coalition for Advancing Research Assessment. (2025). *Working group on ethics and research integrity policy in responsible research assessment for data and ai*. Retrieved July 12, 2026, from https://www.coara.org/working-groups/wg-erip/

COMETH. (2024). *Opinion on the use of generative AI at Paris Brain Institute* [Debate held 8 October 2024]. Retrieved July 13, 2026, from https://myicm.icm-institute.org/system/files/news/cometh_opinion_on_the_use_of_generative_ai_at_paris_brain_institute.pdf

Commission Nationale de l'Informatique et des Libertés. (2025). *Artificial intelligence and the GDPR: CNIL recommendations*. Retrieved July 12, 2026, from https://www.cnil.fr/en/topics/artificial-intelligence-ai

Commission Nationale de l'Informatique et des Libertés. (2026). *AI system development: CNIL's recommendations to comply with the GDPR* [Published 5 January 2026; a step-by-step how-to-sheet series (cnil.fr/fr/ai-how-to-sheets) and a compliance checklist (cnil.fr/sites/default/files/2026-01/ai_checklist.pdf) accompany the main recommendations]. Retrieved July 14, 2026, from https://www.cnil.fr/en/ai-system-development-cnils-recommendations-to-comply-gdpr

Committee on Publication Ethics. (2023). *Authorship and ai tools: COPE position statement*. Retrieved July 12, 2026, from https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Corvol, J. C., Lau, B., Schmidt, L., Sitt, J., Aymé, S., Antunes, M., Whitmarsh, S., Zujovic, V., Chaillou, S., Lesaulnier, F., Gibert, M., & Levy-Marchal, C. (2024). *Data policy: Paris brain institute*. Retrieved July 14, 2026, from https://parisbraininstitute.org/media/410/download

Council for National Open Science Coordination. (2025). *CoNOSC addresses OS impact, AI, and resiliency of open science policies* [Meeting held in Ljubljana, 26–27 June 2025]. Retrieved July 14, 2026, from https://conosc.org/2025/10/21/conosc-addresses-os-impact-ai-and-resiliency-of-open-science-policies/

Council of Europe. (2024). *Framework convention on artificial intelligence and human rights, democracy and the rule of law (cets no. 225)*. Retrieved July 12, 2026, from https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Elsevier. (2026a). *The use of generative ai and ai-assisted technologies in the review process*. Retrieved July 12, 2026, from https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-the-review-process

Elsevier. (2026b). *The use of generative ai and ai-assisted technologies in writing for elsevier*. Retrieved July 12, 2026, from https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier

European Commission. (2026). *European health data space regulation* [Explainer page; most secondary-use obligations stated to apply from March 2029]. Retrieved July 13, 2026, from https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en

European Commission & European Research Area Forum. (2026). *Living guidelines on the responsible use of generative ai in research* (First published March 2024; updated 8 May 2026). European Commission, Directorate-General for Research and Innovation. Retrieved July 12, 2026, from https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/updated-era-living-guidelines-responsible-use-generative-ai-research-2026-05-08_en

European Data Protection Board. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of ai models*. Retrieved July 12, 2026, from https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en

European Parliament and Council of the European Union. (2016). *Regulation (eu) 2016/679 (general data protection regulation)*. Retrieved July 12, 2026, from https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament and Council of the European Union. (2024). *Regulation (eu) 2024/1689 laying down harmonised rules on artificial intelligence (artificial intelligence act)*. Retrieved July 12, 2026, from https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Parliament and Council of the European Union. (2025). *Regulation (eu) 2025/327 establishing the european health data space*. Retrieved July 12, 2026, from https://eur-lex.europa.eu/eli/reg/2025/327/oj

European Research Council. (2026). *Erc clarifies limits of ai use in grant evaluation*. Retrieved July 12, 2026, from https://erc.europa.eu/news-events/news/erc-clarifies-limits-ai-use-grant-evaluation

Freshfields. (2026). *Eu ai act unpacked #34: The final digital omnibus on ai* [Analysis of the amending regulation postponing the AI Act's high-risk obligations, approved by the European Parliament on 16 June 2026 and the Council on 29 June 2026, in force since July 2026]. Retrieved July 27,

2026, from https://www.freshfields.com/en/our-thinking/blogs/technology-quotient/eu-ai-act-unpacked-34-the-final-digital-omnibus-on-ai-key-amendments-to-the-a-102nber

Gerlich, M. (2025). Ai tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Humlum, A., & Vestergaard, E. (2025). *Large language models, small labor market effects* (tech. rep. No. 33777). National Bureau of Economic Research. Retrieved July 27, 2026, from https://www.nber.org/papers/w33777

Inria. (2026). *Agentic AI: Deployment, adoption and impacts*. Retrieved July 13, 2026, from https://www.entreprises.gouv.fr/files/files/Actualites/2026/g7/agentic-ai-inria-short.pdf

Inserm. (2025). *Guide de bonnes pratiques de l'IA à l'Inserm* (February 2025). Institut national de la santé et de la recherche médicale. Retrieved July 12, 2026, from https://www.ofis-france.fr/wp-content/uploads/2025/04/Guide-de-bonnes-pratiques-de-lIA-a-lInserm-fev2025-vf.pdf

Inserm. (2026). *Recommendations on the use of AI systems at Inserm* (March 2026). Institut national de la santé et de la recherche médicale. Retrieved July 12, 2026, from https://lorier.inserm.fr/wp-content/uploads/2026/03/Recommendations-on-the-Use-of-AI-Systems-at-Inserm.pdf

International Committee of Medical Journal Editors. (2024). *Recommendations: Use of artificial intelligence in the conduct and reporting of research*. Retrieved July 12, 2026, from https://www.icmje.org/recommendations/browse/artificial-intelligence/ai-use-by-authors.html

International Energy Agency. (2025). *Energy and ai*. International Energy Agency. Retrieved July 12, 2026, from https://www.iea.org/reports/energy-and-ai/executive-summary

ISO 9001. (2015). *ISO 9001:2015: Quality management systems — requirements*. Retrieved July 14, 2026, from https://www.iso.org/standard/62085.html

ISO/IEC 42001. (2023). *ISO/IEC 42001:2023: Information technology — artificial intelligence — management system*. Retrieved July 14, 2026, from https://www.iso.org/standard/42001

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, *4*(9), 100804. https://doi.org/10.1016/j.patter.2023.100804

Lane, M., Williams, M., & Broecke, S. (2023). *The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers* (tech. rep. No. 288). OECD Publishing. https://doi.org/10.1787/ea0a0fe1-en

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative ai on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3706598.3713778

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). Gpt detectors are biased against non-native english writers. *Patterns*, *4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Liu, J., Xia, C. S., Wang, Y., & Zhang, L. (2023). Is your code generated by ChatGPT really correct? rigorous evaluation of large language models for code generation. *Advances in Neural Information Processing Systems (NeurIPS)*, *36*. Retrieved July 12, 2026, from https://proceedings.neurips.cc/paper_files/paper/2023/hash/43e9d647ccd3e4b7b5baab53f0368686-Abstract-Conference.html

Mistral AI. (2025). *Our contribution to a global environmental standard for ai* [Lifecycle analysis of Mistral Large 2 conducted with Carbone 4 and ADEME, peer reviewed by Resilio and Hubblo, following the AFNOR Frugal AI methodology and ISO 14040/44]. Retrieved July 27, 2026, from https://mistral.ai/news/our-contribution-to-a-global-environmental-standard-for-ai

Naddaf, M., & Quill, E. (2026). *Hallucinated citations are polluting the scientific literature. what can be done?* Retrieved July 12, 2026, from https://www.nature.com/articles/d41586-026-00969-z

National Institutes of Health. (2023). *The use of generative artificial intelligence technologies is prohibited for the nih peer review process (not-od-23-149)*. Retrieved July 12, 2026, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-149.html

National Institutes of Health. (2025a). *Protecting human genomic data when developing generative artificial intelligence tools and applications (not-od-25-081)*. Retrieved July 24, 2026, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-081.html

National Institutes of Health. (2025b). *Responsibly developing and sharing generative artificial intelligence tools using NIH controlled-access data (not-od-25-118)*. Retrieved July 13, 2026, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-118.html

National Institutes of Health. (2025c). *Supporting fairness and originality in nih research applications: Policy on ai use and submission limits*. Retrieved July 12, 2026, from https://grants.nih.gov/news-events/nih-extramural-nexus-news/2025/07/apply-responsibly-policy-on-ai-use-in-nih-research-applications-and-limiting-submissions-per-pi

OECD. (2023). *Oecd employment outlook 2023: Artificial intelligence and the labour market*. OECD Publishing. Retrieved July 12, 2026, from https://www.oecd.org/en/publications/oecd-employment-outlook-2023_08785bba-en.html

OECD. (2024). *Recommendation of the council on artificial intelligence (oecd/legal/0449)* [Adopted 2019, updated 2024]. Retrieved July 12, 2026, from https://oecd.ai/en/ai-principles

Paris Brain Institute. (2020). *Our commitments: Ethics and professional conduct charter and data governance policy*. Retrieved July 12, 2026, from https://parisbraininstitute.org/our-commitments

Paris Brain Institute, RSSI. (2024). *Information security guide for staff* [Version 1, 12 December 2024; internal document, ICM Dokmee, access restricted to ICM staff; also available in French]. Retrieved July 14, 2026, from https://icm.dokmee.com/DokmeeECM/%5C#/viewer/eyJjYWJpbmV0aWQiOiJhYmJlNjlxNS0yNDFmLTRiYmQtYTY3Mi0zMjI3NDhlZDY4NWIiLCJmaWxlaWQiOiIyY2M2N2I0My05ZDg1LTQwOWQtOGZhZC1iOTFhMzNlNWUxZWMifQ%5C%3D%5C%3D

Paris Brain Institute, RSSI. (2024). *Rssi-ref-010: Politique de sécurité du développement logiciel* [Version 1.0, 9 October 2024; internal document, ICM Dokmee, access restricted to ICM staff]. Retrieved July 14, 2026, from https://icm.dokmee.com/DokmeeECM/%5C#/viewer/eyJjYWJpbmV0aWQiOiJhYmJlNjlxNS0yNDFmLTRiYmQtYTY3Mi0zMjI3NDhlZDY4NWIiLCJmaWxlaWQiOiI3MzNkYTFhZC0zNDY4LTQ3YWEtOGJhOS1hZjAzM2NlOWY1YTYifQ%5C%3D%5C%3D

Paris Brain Institute, RSSI. (2026). *Rssi-ref-004: Charte informatique* [Version 2.0, March 2026; internal document, ICM Dokmee, access restricted to ICM staff]. Retrieved July 14, 2026, from https://icm.dokmee.com/DokmeeECM/%5C#/viewer/eyJjYWJpbmV0aWQiOiJhYmJlNjlxNS0yNDFmLTRiYmQtYTY3Mi0zMjI3NDhlZDY4NWIiLCJmaWxlaWQiOiI4Y2YwNTA1ZS1lZTQ5LTRlYmQtYTIwNC1iNWJkMzRlNjU2MTIifQ%5C%3D%5C%3D

Paris Brain Institute, RSSI. (under review). *Rssi-doc-003: Guide de bonnes pratiques de sécurité de l'utilisation de l'intelligence artificielle* [Draft, not yet published; internal document, access restricted to ICM staff; add the ICM Dokmee URL once the document is formalised].

Paris Brain Institute, RSSI. (in preparation). *Rssi-doc-00x: Procédure aipd et gestion du risque lié à l'intelligence artificielle* [Reference number RSSI-DOC-00X is a placeholder; internal document in preparation and validation, not yet published; access to be restricted to ICM staff].

Paris Brain Institute, RSSI. (in preparation). *Rssi-ref-00x: Charte informatique de l'utilisation de l'intelligence artificielle* [Reference number RSSI-REF-00X is a placeholder; internal document in preparation and validation, not yet published; access to be restricted to ICM staff].

Paris Brain Institute, RSSI. (in preparation). *Rssi-ref-015: Politique de sécurité de l'utilisation de l'intelligence artificielle* [Internal document in preparation and validation, not yet published; access to be restricted to ICM staff].

Perrigo, B. (2023). *Exclusive: OpenAI used kenyan workers on less than $2 per hour to make ChatGPT less toxic* [*TIME*, 18 January 2023]. Retrieved July 13, 2026, from https://time.com/6247678/openai-chatgpt-kenya-workers/

Public Library of Science. (2026). *Ethical publishing practice: Ai tools*. Retrieved July 12, 2026, from https://journals.plos.org/plosone/s/ethical-publishing-practice

Romeo, G., & Conti, D. (2025). Exploring automation bias in human–ai collaboration: A review and implications for explainable AI [Published online July 2025, assigned to Vol. 41, No. 1 (2026)]. *AI & Society*, *41*(1), 259–278. https://doi.org/10.1007/s00146-025-02422-7

Sorbonne Université, Faculté de Santé, Commission de déontologie et d'Intégrité Scientifique. (2026). *Recommandations d'utilisation de l'intelligence artificielle générative dans le cadre de la recherche* [April 2026]. Retrieved July 24, 2026, from https://sante.sorbonne-universite.fr/actualites/recommandations-dutilisation-de-lintelligence-artificielle-generative-dans-le-cadre-de

Spracklen, J., Wijewickrama, R., Sakib, A. H. M. N., Maiti, A., Viswanath, B., & Jadliwala, M. (2025). We have a package for you! a comprehensive analysis of package hallucinations by code generating LLMs. *Proceedings of the 34th USENIX Security Symposium*. Retrieved July 13, 2026, from https://www.usenix.org/conference/usenixsecurity25/presentation/spracklen

Springer Nature. (2026). *Artificial intelligence (ai): Nature portfolio editorial policies*. Retrieved July 12, 2026, from https://www.nature.com/nature-portfolio/editorial-policies/ai

Stanford University. (2025). *Responsible ai at stanford*. Retrieved July 12, 2026, from https://uit.stanford.edu/security/responsibleai

Swiss Young Academy. (2025). *Impact of ai on early career researchers: Challenges, opportunities and responsibilities*. Swiss Young Academy. Retrieved July 12, 2026, from https://swissyoungacademy.ch

UK Biobank. (2026). *Use of artificial intelligence (ai) applications and models*. Retrieved July 13, 2026, from https://community.ukbiobank.ac.uk/hc/en-gb/articles/21922841787293-Use-of-Artificial-Intelligence-AI-applications-and-models

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. Retrieved July 12, 2026, from https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Upwork Research Institute. (2024). *Upwork study finds employee workloads rising despite increased c-suite investment in artificial intelligence* [Survey of 2,500 workers and executives in the United States, United Kingdom, Australia, and Canada, April–May 2024]. Retrieved July 27, 2026, from https://investors.upwork.com/news-releases/news-release-details/upwork-study-finds-employee-workloads-rising-despite-increased-c

U.S. Department of Justice. (2018). *CLOUD Act resources*. Retrieved July 24, 2026, from https://www.justice.gov/criminal/criminal-oia/cloud-act-resources

Wellcome Trust. (2023). *Joint statement on the responsible use of generative ai in research (research funders policy group)*. Retrieved July 12, 2026, from https://wellcome.org/about-us/positions-and-statements/joint-statement-generative-ai

World Association of Medical Editors. (2023). *Chatbots, generative ai, and scholarly manuscripts: WAME recommendations on chatgpt and chatbots*. Retrieved July 12, 2026, from https://wame.org/page3.php?id=110

E About this document

During the preparation of this document, the author(s) used Claude Sonnet 5 (Anthropic) in July 2026 to source information and improve the language and readability of the text. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for its content.

Release Date	Version No.	Revision History	Drafted by	Reviewed by	Validated by	Approved by
15/07/2026	0.1	First draft.	Stephen Whitmarsh	CISO/RSSI	n/a	n/a
24/07/2026	0.2	Integrated feedback, SU policy and CLUSIF	Stephen Whitmarsh	n/a	CISO/RSSI	n/a
27/07/2026	0.3	Integrated feedback, pending DSI documents, sent for review to team	Stephen Whitmarsh	n/a	n/a	n/a
31/07/2026	0.4	Integrated feedback from team, simplified introduction	Stephen Whitmarsh	n/a	n/a	n/a