

State of AI confidentiality 2026

Lawyers, accountants, HR teams, banks and hospitals now put client and employee files into Claude, ChatGPT, Gemini, Copilot and Le Chat. This report sets out what the law and the regulators ask of them, how the AI tools audited by the French bar answer, how well today's tools detect French and Moroccan personal data, and what a firm can do about it this year.

Hidden Tokens research · edition 1 · published 25 October 2026 · facts checked between 24 and 25 October 2026

14/15

legal AI tools in the French bar's 2025 audit name a US company among those hosting the LLM.

CNB grid, June 2025 ¹

9/15

declare nothing on pseudonymisation or anonymisation ("N/C" or "N/A").

CNB grid, June 2025 ¹

77.5%

of French and Moroccan direct identifiers caught by Presidio's default recognisers on HT-Bench.

HT-Bench 0.2.0, 25 Oct. 2026
¹¹

KEY FINDINGS

- **The rules already exist; AI just makes them urgent.** Professional secrecy, the GDPR and Moroccan law 09-08 apply to a prompt exactly as they apply to an email. The EU AI Act adds AI-specific duties, but it does not replace any of them.
- **The French bar names the problem plainly.** Its reading grid excludes, as a sovereignty criterion, hosting by companies subject to extraterritorial requisition laws, and gives US companies as the example. Fourteen of the fifteen tools it summarises name a US company among those hosting the language model.
- **Pseudonymisation is the missing field.** Only five of the fifteen summary forms declare pseudonymisation (3) or anonymisation (2). The bar's April 2026 ethics guide recommends pseudonymising or anonymising data before using a generative AI tool.
- **Generic detectors miss local identifiers.** On 366 synthetic French and Moroccan documents, Presidio's defaults caught 77.5% of direct identifiers, including 69.8% of Moroccan ID card numbers, 5.5% of Moroccan bank account numbers (RIB) and 1.8% of company names. A tool built for these formats (Hidden Tokens 2.28.0) is at $\geq 99\%$ in every pack on HT-Bench (100% on the 2.28.0 test split). That is a synthetic benchmark run by the tool's own publisher, not a claim about real documents.
- **No single category solves it.** Enterprise AI plans govern what the provider does with data it receives in clear; DLP blocks or strips data; redaction APIs need engineering. Reversible pseudonymisation on the workstation is the newest category and the least measured, ours included.

1 Why now

In 2025, 20.0% of EU enterprises with ten or more employees used artificial intelligence, up from 13.5% a year earlier, and 8.8% used it to generate written or spoken language³. In France, 18% of companies used at least one AI technology in 2025, eight points more than in 2024; 43% of those used generative text or speech tools⁴. Among the tools staff reach for are general-purpose assistants: Claude, ChatGPT, Gemini, Microsoft Copilot and Mistral's assistant (Le Chat, renamed Vibe in August 2026).

Every one of these assistants is, by default, a cloud service: the prompt and every attached file travel to the provider's servers, where the model reads them in clear. Business plans add contractual commitments (no training on customer data by default, retention controls, sometimes a choice of data region), but they do not change that basic fact. For professions bound by secrecy, the question is not whether the provider is trustworthy. It is whether the firm may hand the information over at all.

Professional secrecy

In France, the lawyer's professional secrecy covers, "in all matters", consultations, correspondence, meeting notes and every document in the file (law no. 71-1130 of 31 December 1971, article 66-5)⁹; breaching a secret held by profession is a criminal offence punishable by one year's imprisonment and a €15,000 fine (Criminal Code, article 226-13)¹⁰. Morocco's Criminal Code punishes the same breach (article 446). Secrecy belongs to the client, not to the firm: a firm cannot waive it by accepting a provider's terms of service.

GDPR

A prompt that contains names, identity numbers or the facts of an identifiable person's case is personal data. Sending it to an AI provider makes the provider a processor or a recipient (article 28), and may be a transfer outside the EU (chapter V). The GDPR defines pseudonymisation as processing after which data "can no longer be attributed to a specific data subject without the use of additional information" kept separately (article 4(5)); pseudonymised data remain personal data for whoever can re-identify them (recital 26), and articles 25 and 32 cite pseudonymisation as a protection measure⁵. The EDPB's guidelines 01/2025 on pseudonymisation were adopted on 16 January 2025 as a version for public consultation; no final version had been published by October 2026⁶. In *EDPS v SRB* (4 September 2025), the Court of Justice held that pseudonymised data are not personal data "in all cases and for every person": for a recipient that cannot reasonably re-identify them, they may not be, while the controller's own duties are assessed from its own point of view⁷. In practice, pseudonymisation does not take a firm out of the GDPR. It can reduce what the AI provider receives to data it cannot attribute to anyone, provided the key stays with the firm.

Morocco: law 09-08

Law no. 09-08 of 18 February 2009 governs the processing of personal data in Morocco and created the national control commission, the CNDP (article 27)⁸. It applies to processing carried out through an AI tool as to any other. Article 43 allows a transfer of personal data to a foreign state only if that state ensures a sufficient level of protection; article 44 lists the exceptions, including the person's express consent and an express, reasoned authorisation from the CNDP⁸. Most general-purpose AI assistants process prompts outside Morocco, so sending client data to them raises the transfer question first. The

implementing decree already asks that data passed to a third party for statistical or research purposes be "rendered anonymous or coded" (decree no. 2-09-165, article 31)⁸. We found no CNDP deliberation specific to generative AI in 2025–2026, and could not reach the CNDP's website to check its latest publications.

The EU AI Act timeline

Regulation (EU) 2024/1689 entered into force on 1 August 2024. Its general provisions (including the duty of AI literacy) and its prohibited practices have applied since 2 February 2025, the rules on general-purpose AI models since 2 August 2025, and the regulation as a whole since 2 August 2026¹³. The Digital Omnibus on AI, Regulation (EU) 2026/1744 (published on 24 July 2026, in force since 27 July 2026), postponed the high-risk obligations: to 2 December 2027 for the systems listed in Annex III and to 2 August 2028 for those covered by Annex I¹⁴. Annex III includes, among others, AI used in employment decisions, which concerns HR teams. For a firm that only uses a general-purpose assistant on client files, confidentiality is still governed by secrecy and data protection law, not by the AI Act. The GDPR part of the Digital Omnibus (COM(2025) 837), which would amend the definition of personal data and add a provision on pseudonymisation, was still being negotiated in October 2026¹⁵.

2 What regulators ask: the French bar's audit of 15 legal AI tools

The most concrete text any regulator of a profession has published on the subject comes from France's National Bar Council (Conseil national des barreaux, CNB). Its AI working group's *Grille de lecture* (1st edition, June 2025, v1.2)¹ does two things. It sets out the criteria a firm should use to choose a legal AI tool, and it sums up fifteen tools it consulted in a ten-field summary form ("fiche de synthèse"): where the LLM is hosted and by which company, where client data is hosted and by which company, encryption, secure RAG, reuse of client data by the LLM, pseudonymisation or anonymisation, which of the vendor's staff can access client data, and other measures.

The sovereignty criteria

The grid ranks four data-sovereignty criteria, in order of priority:

1. where client data is hosted (France, or else the European Union);
2. the nationality of the company that owns the servers hosting client data (French, or else European), **excluding any company whose national law contains extraterritorial provisions allowing its authorities to requisition data stored on servers owned by companies of that nationality, "such as US companies under their current legislation"**;
3. where the large language model is hosted (France, or else the EU);
4. the nationality of the company that owns the servers hosting the LLM (same rule as criterion 2).

It adds a prerequisite: the tool must not reuse data to train the model. All tools audited say they apply it, but the grid notes that such a statement from a proprietary model is "in general neither provable nor verifiable". To remove these risks, it names two routes: full prior anonymisation of the data, or installing the LLM on the firm's own systems. The CNB's ethics guide on AI (April 2026)² widens the first route to pseudonymisation.

What the fifteen summary forms declare

We read the fifteen forms in the PDF published by the CNB and transcribed two fields. The wording is the CNB's, translated; "N/C" means not communicated, "N/A" not applicable.

Tool	Company hosting the LLM	Pseudonymisation / anonymisation
Doctrine	US (Microsoft)	Pseudonymisation
Lefebvre Dalloz	US (Microsoft)	N/C
Jimini	French (Mistral) and US (Microsoft)	Pseudonymisation
Juri'Predis	US (Microsoft)	Pseudonymisation
Jus Mundi	US (Microsoft)	Anonymisation
Lamy Liaisons	US (Microsoft)	N/C

Legora	US (Microsoft, AWS and Amazon Bedrock)	N/C
Lexbase	US (Microsoft) and on premises at Lexbase	Anonymisation
LexisNexis	US (Microsoft, AWS and Amazon Bedrock)	No, but deletion after a few minutes or days
Notae	US (Google)	N/C
Ordalie	US (Microsoft) and on premises at Ordalie	N/C
Septeo	US (Microsoft)	N/C
Cedie	N/A (runs on the user's machine)	N/A
Predictice	US	N/C (may be specified in its Trust Center)
Haiku	US	N/C

Source: CNB, *Grille de lecture*, June 2025, v1.2, summary forms (pp. 29–43 of the PDF). Our transcription and count, 24–25 October 2026. Tools are listed in the grid's order.

- **LLM hosting.** 14 of 15 forms name a US company among those hosting the LLM: 11 name only US companies, 3 name a US company alongside a French model or an on-premises installation (Jimini, Lexbase, Ordalie). The fifteenth, Cedie, runs on the user's machine.
- **Pseudonymisation or anonymisation.** 5 forms declare a mechanism (3 pseudonymisation, 2 anonymisation), 1 answers "no" with short retention instead (LexisNexis), 8 read "N/C" and 1 "N/A". In total, **9 of 15 declare nothing** in that field.

This is not a ranking. The forms reflect what vendors declared to the working group in 2025 (the grid says so: "subject to change and depending on what the companies declared, found and/or sent"), and several vendors have published more since. It is a snapshot of where a careful regulator found the market: almost every tool depends on a US cloud for the model, and most do not say whether the model receives names in clear.

Beyond the bar

Other regulators have said less, and less precisely. The EDPB's pseudonymisation guidelines are still a consultation draft, and the CNDP has published no deliberation specific to generative AI that we could find (section 1). The bar's own rules set the bar high: the National Internal Regulation describes professional secrecy as "general, absolute and unlimited in time" (article 2.1)¹⁶, and the CNB's April 2026 guide recommends to "systematically pseudonymise or anonymise data before any use of a generative AI tool", while warning that pseudonymisation does not fit every task and that people can be

re-identified by cross-referencing². The ten-field form is the most detailed checklist any regulator of a profession has published, and it is useful well beyond lawyers.

3 How well tools detect French and Moroccan personal data

Pseudonymisation is only as good as the detection behind it: a name the tool misses reaches the AI in clear. Most public benchmarks of PII detection are in English, with US identifiers. We built HT-Bench to measure the identifiers found in French and Moroccan files: CIN (Moroccan ID card), ICE, IF and RC (Moroccan company identifiers), CNSS, RIB, French NIR, SIREN and SIRET, IBANs, phone numbers, and personal names from the Maghreb and France.

Results by industry pack

HT-Bench 0.2.0, held-out test split: 366 synthetic documents, 3,420 labelled values of which 2,523 are direct identifiers, and 3,029 decoys (amounts, dates, law articles, ordinary words) that must stay readable. Hidden Tokens connector 2.28.0 against Microsoft Presidio 2.2.358 with its default recognisers, run on 25 October 2026 on the same documents and scored by the same code¹¹. The 2.27.0 column is the previous run on the same test split.

Pack	Docs	Hidden Tokens 2.28.0, direct identifiers	2.27.0	Hidden Tokens, precision	Presidio, direct identifiers	Presidio, precision
telecom-ma Telecom (Morocco)	36	100.0% (242/242)	100.0%	82.7%	81.8% (198/242)	51.8%
core Core (all documents)	42	100.0% (322/322)	99.7%	93.5%	87.0% (280/322)	59.9%
banque-assurance Banking and insurance	96	100.0% (652/652)	99.5%	81.7%	64.4% (420/652)	54.1%
expertise-comptable Accounting	84	100.0% (599/599)	99.3%	95.5%	81.8% (490/599)	60.6%
sante Healthcare	36	100.0% (204/204)	96.1%	82.7%	86.3% (176/204)	50.5%
juridique-fr Legal (France)	36	100.0% (276/276)	96.0%	87.2%	82.2% (227/276)	50.3%
rh-paie HR and payroll	36	100.0% (228/228)	92.1%	89.3%	72.4% (165/228)	51.8%
All packs	366	100.0% (2,523/2,523)	98.2%	87.4%	77.5% (1,956/2,523)	54.7%

Recall on direct identifiers; a value counts as detected only if every letter and digit of it is coded. In red: 2.27.0 packs below the 99% per-pack target Hidden Tokens set itself. Source: [ht-bench-2.28.0.json](#) (2.27.0: [ht-bench-2.27.0.json](#)).

In short: Hidden Tokens 2.28.0 is at $\geq 99\%$ of direct identifiers in every pack on HT-Bench (100% on the 2.28.0 test split, 2,523/2,523), with a precision of 87.4% and 5.8% of decoys coded. Presidio's defaults detect 77.5% of the same values, with a precision of 54.7% and 37.6% of decoys coded. This is a synthetic benchmark run by the tool's own publisher: it is not a claim about real documents.

What generic detectors miss

- **Local formats.** Presidio's default recognisers include none for Moroccan formats. It caught 69.8% of CIN numbers (134/192), 66.0% of ICE numbers (105/159) and 5.5% of RIBs (6/109), mostly through recognisers written for other countries (US driver's licence, Indian PAN, US bank number). The same recognisers flag almost any long number, which is why Presidio coded 37.6% of the decoys.
- **Company names.** Presidio's defaults do not tag organisations: 1.8% of company names detected (5/279). In a commercial file, the counterparty's name often identifies the client.
- **French social security numbers.** Presidio caught 32.5% of NIRs (13/40). Written with spaces or a dash before the key, they were also Hidden Tokens' weakest format in 2.27.0 (26/40); 2.28.0 detects all 40.
- **What Hidden Tokens still misses** is outside the direct identifiers: postal addresses 92.4% (170/184), dates of birth 79.1% (68/86) and cheque numbers 75.0% (36/48).

Presidio is a toolkit meant to be extended with custom recognisers. This comparison measures its defaults against a finished product, not what an engineering team could build on Presidio. It does show what a firm gets if it deploys a generic detector on French or Moroccan files without that work.

Method and limits

- **Synthetic data only.** Names are common Moroccan, Maghrebi and French names recombined at random, in Latin script; numbers are random with valid check digits. No real file was used.
- **Held-out split, with a caveat.** The engine is tuned on a separate development split; the published numbers come from the test split (seed `ht-bench-v0-test`, SHA-256 `4881c8d2...e10b`). The 2.28.0 fixes were made on the development split, but the 2.27.0 test gaps had been published before, and both splits are built from the same document templates. The test split shows the fixes hold on values and phrasings the engine was not tuned on; it does not show how the engine handles formats nobody has written a template for.
- **Not independent.** HT-Bench is Hidden Tokens' own benchmark, run by its publisher. Code, data generator and raw results are public so that anyone can re-run it; it has not been audited by a third party.
- **Not covered:** Arabic script, scanned documents and OCR errors, Excel and Word layouts (covered by a separate leak test), and the notary pack `notariat-ma`, which ships with the connector but is not measured: HT-Bench 0.2.0 has no notary documents.

Full method, per-type tables and how to reproduce: hiddentokens.com/benchmark¹².

4 Market map: four ways to keep client data out of an AI

Four categories of product answer the question "how do we use AI on confidential files?". They are often sold as substitutes; they are better read as layers. What separates them is what the AI finally receives, and who holds the means to read it.

Category	What the AI receives	Values put back in the answer	Who runs it	Examples (checked 25 Oct. 2026)
1. Enterprise AI plans	The data in clear, under contract: no training by default, retention controls, sometimes a data region	Not applicable	The AI provider; the firm's admin sets options	ChatGPT Enterprise and Business ¹⁷ , Claude Team and Enterprise ¹⁸ , Microsoft Copilot ¹⁹ , Gemini for Workspace ²⁰ , Mistral Vibe ²¹
2. DLP and CASB for generative AI	Nothing, if the prompt is blocked; the prompt minus what was stripped, if redacted	Generally no	IT and security teams, across the organisation	Microsoft Purview (warn or block) ²² , Nightfall (redact or block) ²³ , Strac (also announces reversible pseudonymisation) ²⁴
3. PII detection and redaction APIs	Whatever the pipeline built on them sends	Possible, if engineered	Developers, inside the firm's own applications	Presidio (open source; <code>encrypt / decrypt</code> operators) ²⁵ , Limina, formerly Private AI (containers; re-identification API) ²⁶
4. Reversible pseudonymisation layers	Consistent codes or placeholders instead of names and identifiers	Yes, automatically	A proxy, a vault service or the workstation, depending on the product	LLM Guard (Vault + Deanonymize) ²⁷ , LiteLLM's Presidio guardrail ²⁸ , Skyflow LLM Privacy Vault ²⁹ , Protecto ³⁰ , Hidden Tokens (workstation)

Categories and examples are ours. Each example is described from its vendor's public documentation, read on 25 October 2026; the list is not exhaustive.

Enterprise plans: what the vendors' own pages say

- **OpenAI** (page updated 8 January 2026): no training on business data by default; workspace admins control retention; deleted conversations are removed within 30 days. Data residency at rest is offered for Enterprise, Edu, Healthcare and the API; ChatGPT Business is not listed¹⁷.
- **Anthropic** (pages updated June–August 2026): no training on commercial products' inputs and outputs by default; Enterprise offers custom retention, with a 30-day minimum; zero data retention applies only to eligible APIs and Claude Code, not to the claude.ai chat; data is stored in the US¹⁸.
- **Microsoft Copilot**, formerly Microsoft 365 Copilot (page updated 8 October 2026): prompts and responses are not used to train foundation models; the service follows the EU Data Boundary, except that models provided by Anthropic are currently excluded from it¹⁹.
- **Google Gemini for Workspace** (updated 6 October 2026): no training on customer data without prior permission; admins set retention from 90 days to indefinite; processing can be limited to the US or the EU²⁰.
- **Mistral**, whose Le Chat became Vibe on 12 August 2026: Enterprise customers are opted out of training by default; data is hosted in the EU by default; zero data retention is not available for the chat, whatever the plan²¹.

Every plan in this category improves what happens to the data *after* the provider receives it. None changes the fact that it receives it in clear. That is why the bar's grid asks a separate question about pseudonymisation, and why the other three categories exist.

What to watch

- **Where the key lives.** In category 4, a vault run by a vendor and a key on the workstation are very different answers to the CNB's criteria 2 and 4.
- **Detection quality on your languages.** Presidio's documentation states that, in its default configuration, it contains recognisers and models for English²⁵; section 3 shows what that means on French and Moroccan files.
- **Names change fast.** In 2026 alone, Private AI became Limina (March), Le Chat became Vibe (August), Microsoft 365 Copilot became Microsoft Copilot, and Presidio's documentation moved to a community-owned project²⁵²⁶. Date every claim you rely on.

5 Recommendations: a 10-point checklist

For a law firm, an accounting firm, an HR department or any company putting client or employee files into AI. None of these points requires a particular product.

- 1 **Map real usage first.** Ask teams which AI tools they use, on which files, from which devices. Usage you have not mapped is usage you cannot govern.

- 2 **Classify what may never leave in clear.** Client identities, identity and account numbers, health data, salaries, case numbers. Write the list down; it becomes the configuration of whatever tool you choose.

- 3 **Choose a business plan, never a personal account,** for every AI tool you allow: training on your data off, retention settings known, admin controls on.

- 4 **Ask each vendor the CNB's ten questions,** whatever your profession: who hosts the model and the data, under which jurisdiction, with what encryption, whether data is reused, and whether names are pseudonymised before the model sees them. "N/C" is an answer too.

- 5 **Pseudonymise upstream of the AI where you can,** as the CNB's 2026 guide recommends: the model receives consistent codes, the key stays with you, and the real values are put back afterwards.

- 6 **Keep the key yourself.** Pseudonymisation protects only if whoever receives the coded data cannot re-identify it. A key held by the AI provider or the tool vendor protects little.

- 7 **Test detection on your own documents.** Take ten typical files, run them through the tool, and look for every name and number that survived. Ask the vendor for per-format recall, on your languages and identifiers.

- 8 **Close the side doors.** Text typed straight into a chat, files dragged into a browser, sync folders, browser extensions and agents with file access. Decide which route is authorised and block or monitor the others.

- 9 **Write it down.** An AI policy, a record of processing, a data protection impact assessment where the GDPR or law 09-08 requires one, and the client information your professional rules call for.

- 10 **Keep proof.** Be able to show a client, a DPO or a bar what the AI actually received, file by file, and review it at least once a year.

6 Where Hidden Tokens fits (disclosure)

This report is published by Tardigrade SARL, which sells Hidden Tokens. Hidden Tokens belongs to the fourth category above: desktop software that codes names, identifiers and sensitive amounts on the user's computer before Claude, ChatGPT, Gemini or Copilot reads them, keeps the key on that computer or the organisation's file share, and puts the real values back into the AI's Word or Excel output. Section 3 uses its benchmark; read those numbers as a vendor's own measurement.

What it does not do: it protects only what goes through it (text typed into a chat without its browser extension is not covered), it flags indirect identifiers rather than rewriting all of them, and it supports compliance without certifying it. Its answers to the CNB's ten fields, with their limits, are in the [CNB compliance file](#).

7 Method, conflicts of interest, how to cite

- **Sources.** Primary sources only: legislation, regulators' publications, court decisions and vendors' own documentation. Each fact about a third-party product is dated; products change, and we will correct this page if told at info@hiddentokens.com.
- **The CNB figures** are our own count from the CNB's published PDF, not figures published by the CNB.
- **The detection figures** come from Hidden Tokens' own benchmark (see section 3 for its limits).
- **Not legal advice.** This report describes the law and the market; it does not replace advice on a firm's own situation.

How to cite:

Hidden Tokens, *State of AI confidentiality 2026*, edition 1, 25 October 2026,
<https://hiddentokens.com/state-of-ai-confidentiality-2026/>

The French version, [État de la confidentialité de l'IA 2026](#), says the same thing; the French texts quoted prevail.

Sources

1. CNB AI working group, *Grille de lecture — synthèse des consultations...*, 1st edition, June 2025, v1.2 (French). cnb.avocat.fr/medias/cnb-grille-lecture-intelligence-artificielle-2025-68f775b81ba5b2.65545272.pdf
2. CNB, Rules and practices committee, *La déontologie et l'intelligence artificielle*, 1st edition, April 2026 (French). cnb.avocat.fr/medias/guide-deontologie-ia-6a4cef5ea1d577.28126431.pdf
3. Eurostat, news release of 11 December 2025, and *Use of artificial intelligence in enterprises*, Statistics Explained (2025 data). ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20251211-2
4. INSEE, *Insee Première* no. 2120, July 2026 (French). www.insee.fr/fr/statistiques/9025878
5. Regulation (EU) 2016/679 (GDPR), articles 4(5), 25, 28, 32, chapter V, recital 26. eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

6. EDPB, Guidelines 01/2025 on pseudonymisation, adopted 16 January 2025, version for public consultation. www.edpb.europa.eu/documents/guideline/guidelines-012025-on-pseudonymisation_en
7. CJEU, 4 September 2025, EDPS v SRB, C-413/23 P (press release no. 107/25). curia.europa.eu/site/upload/docs/application/pdf/2025-09/cp250107en.pdf
8. Moroccan law no. 09-08 on the protection of individuals with regard to the processing of personal data (dahir no. 1-09-15 of 18 February 2009) and decree no. 2-09-165, Ministry of Justice Adala portal (French). adala.justice.gov.ma/api/uploads/2024/04/30/Protection%20des%20personnes%20physiques-1714464099884.pdf
9. French law no. 71-1130 of 31 December 1971, article 66-5 (French). www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000006512415
10. French Criminal Code, article 226-13 (French). www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006417945
11. Hidden Tokens, HT-Bench 0.2.0 raw results, test split, connector 2.28.0, 25 October 2026 (JSON); previous run, connector 2.27.0: hiddentokens.com/benchmark/ht-bench-2.27.0.json; hiddentokens.com/benchmark/ht-bench-2.28.0.json
12. Hidden Tokens, *Benchmark: detection per industry pack, and leak test*. hiddentokens.com/benchmark/
13. Regulation (EU) 2024/1689 (AI Act), article 113. eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32024R1689
14. Regulation (EU) 2026/1744 of 8 July 2026 (Digital Omnibus on AI), OJ L of 24 July 2026. eur-lex.europa.eu/eli/reg/2026/1744/oj/eng
15. European Commission, COM(2025) 837, procedure 2025/0360(COD), ongoing. eur-lex.europa.eu/procedure/EN/2025_360
16. CNB, National Internal Regulation of the legal profession (RIN), consolidated 7 July 2026, article 2.1 (French). cnb.avocat.fr/medias/file/cnbrinmaj-08072026-6a4e79014f6b41.49207715.pdf
17. OpenAI, *Enterprise privacy* (updated 8 January 2026) and *Business data*, read 25 October 2026. openai.com/enterprise-privacy/
18. Anthropic Privacy Center: model training (updated 18 August 2026), retention, zero data retention, server location (15 June 2026), read 25 October 2026. privacy.claude.com/en/articles/7996868-is-my-data-used-for-model-training
19. Microsoft Learn, *Data, privacy, and security for Microsoft Copilot* (updated 8 October 2026). learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
20. Google, *Generative AI in Google Workspace Privacy Hub* (updated 6 October 2026). knowledge.workspace.google.com/admin/generative-ai/generative-ai-in-google-workspace-privacy-hub
21. Mistral AI help centre: "Le Chat is now Vibe" (12 August 2026), training opt-out (3 September 2026), zero data retention and data storage (12 August 2026). help.mistral.ai/en/articles/682992-le-chat-is-now-vibe
22. Microsoft Learn, *Microsoft Purview data security and compliance protections for generative AI apps* (27 May 2026). learn.microsoft.com/en-us/purview/ai-microsoft-purview
23. Nightfall AI, "Securely use GenAI" and "Prevent data leakage to shadow AI" solution pages, read 25 October 2026. www.nightfall.ai/solutions/securely-use-genai
24. Strac, "ChatGPT DLP" blog post (1 October 2026). www.strac.io/blog/chatgpt-dlp-data-loss-prevention
25. Presidio documentation (languages, anonymizer); presidio-analyzer 2.2.364 on PyPI (22 July 2026); version measured in section 3: 2.2.358. presidio.dataprivacystack.org/
26. Limina (formerly Private AI), website and documentation (getlimina.ai, docs.getlimina.ai), rebrand press release of 5 March 2026. www.getlimina.ai/en
27. Protect AI, LLM Guard, Anonymize and Deanonimize scanners. protectai.github.io/llm-guard/input_scanners/anonymize/
28. LiteLLM, Presidio PII masking guardrail. docs.litellm.ai/docs/proxy/guardrails/pii_masking_v2
29. Skyflow, *LLM privacy* documentation. docs.skyflow.com/docs/fundamentals/patterns/llm-privacy
30. Protecto, *Privacy Vault* product page. www.protecto.ai/product/privacy-vault/

This report describes the law and the market; it is not legal advice. © 2026 Tardigrade SARL. Claude is a trademark of Anthropic; ChatGPT, Gemini, Copilot, Le Chat and the other products named are trademarks of their respective owners. Hidden Tokens is not affiliated with any of them, nor with the Conseil national des barreaux.