ChatGPT

SOURCES

Topic	Source	Why it matters	Credibility tier
GDPR accountability principle	GDPR Article 5(2) – *Accountability* [1]	Core legal obligation that controllers **demonstrate** compliance with GDPR	Primary (Law)
Controller responsibilities	GDPR Recital 74 – *Responsibility and liability* [2]	Emphasizes duty to implement effective measures and prove their effectiveness	Primary (Law)
Data protection by design/default	GDPR Recital 78 – *Demonstrating compliance via DPbD* [3] [4]	Calls for internal policies & measures (minimization, transparency, etc.) to evidence compliance	Primary (Law)
ChatGPT Taskforce Report (EDPB)	EDPB *ChatGPT Taskforce Report* (23 May 2024) [5] [6]	Joint DPA findings on LLMs; details *documentation expected* (ROPA, DPIA, etc.) and areas of concern (transparency, rights)	Primary (EDPB)
ChatGPT Taskforce Questionnaire	*Annex: ChatGPT TF Questions* (2024) [7] [8]	Explicit list of evidence DPAs asked from an LLM provider (record of processing, DPIA, legal basis, etc.)	Primary (EDPB)
CNIL AI compliance guide	CNIL *AI Development & GDPR Recommendations* (2023–2025) [9] [10]	Practical steps for controllers: roles, legal basis, minimization, transparency, rights, security, DPIA	Primary (DPA)
CNIL AI compliance checklist	CNIL *AI compliance checklist* (2025) [11] [10]	Summarizes GDPR controls for AI: e.g. determine if GDPR applies, define purposes, document Art.14 exemptions, etc.	Primary (DPA)
EDPB Opinion on AI models	EDPB *Opinion on use of personal data in AI models* (Dec 2024) [12] [13]	Confirms GDPR applies to AI; guidance on anonymity tests and legitimate interest for AI use	Primary (EDPB)
European Commission Q&A	EC FAQ: *How to demonstrate GDPR compliance?* [14] [15]	Highlights accountability tools (DPIA, DPO, codes/certifications) and need for records to prove compliance	Primary (EU Commission)

Topic	Source	Why it matters	Credibility tier
ICO AI Auditing Guidance	ICO *AI Auditing Framework* (2020, UK) [16] [17]	Practical audit criteria for AI systems (documentation of processing, lawful basis justification, etc.) – used as best practice	Secondary (DPA)
ICO Accountability Framework	ICO *Accountability Framework* (2021, UK) [18] [19]	Emphasizes maintaining records (ROPA), data mapping, and supply-chain oversight as proof of compliance	Secondary (DPA)
CNIL AI news (July 2025)	CNIL Press Release on AI Recommendations [20] [21]	Notes GDPR's applicability to AI (due to memorization) and need for documented analysis and safeguards in model development	Secondary (DPA)

(At least 10 sources, with at least 6 primary sources above.)

GDPR-GR REPORTING REQUIREMENTS (MINIMAL)

2.1 What must be demonstrable (bullet list):

- **Lawful Basis & Purpose Specification** – **GDPR Anchor**: Art.5(1)(a),(b), Art.6, Art.9. **Evidence expected**: Clear documentation of each processing purpose and its legal basis, including any Legitimate Interests Assessments or consent records. For AI training on personal data, show a purpose compatibility test if re-using data [22] [23]. **Failure modes**: Missing or incorrect legal basis (e.g. relying on legitimate interest without a documented balancing test, or processing special-category data with no Art.9(2) exception), or vague purpose definitions leading to "purpose creep" [24] [25].

- **Record of Processing & Data Mapping** – **GDPR Anchor**: Art.30, Art.5(2). **Evidence expected**: A current Record of Processing Activities (ROPA) covering the chatbot/AI, detailing data categories, sources, subjects, purposes, recipients, retention, and transfers [7]. Data flow maps or inventories showing how personal data enters, moves through, and exits the AI system (including via third-party APIs) [16]. **Failure modes**: No ROPA or an incomplete one (auditors can't tell what data is processed for what), failure to account for personal data in training datasets, or forgetting to include the AI system in enterprise data inventories [26] [27].

- **Transparency & Notice to Individuals** – **GDPR Anchor**: Art.12–14. **Evidence expected**: User-facing privacy notice and disclosures specific to the AI agent, covering its identity, purposes, data uses (including if user inputs will be stored or used to improve models), data retention, and rights [28] [29]. If personal data is obtained indirectly (e.g. web-scraped training data), documented assessment of Art.14(5)(b) "disproportionate effort" exception and evidence of alternative public notice (e.g. a public website privacy statement listing data sources/categories) [30] [31]. **Failure modes**: Opaque AI

services (users not told they are interacting with an AI or how their data is used), privacy policies that omit model training or data sharing information, or unjustified omission of notice to data subjects whose data was scraped (no proof that informing them individually was impossible) [32] [33].

- **Data Subject Rights Readiness** – **GDPR Anchor**: Art.15–22. **Evidence expected**: Documented procedures and technical capabilities to handle access, rectification, erasure, objection, and restriction requests related to data used or produced by the AI [34] [35]. For example, an internal policy or runbook for processing rights requests, logs of past requests (if any) and outcomes, and interface or support process for users to submit requests. If full compliance is technically challenging (e.g. rectifying AI model outputs), evidence of mitigating measures (such as offering deletion if rectification is infeasible [36]). **Failure modes**: No easy way for individuals to exercise their rights (e.g. no contact info or process), or the AI system retains personal data in a model such that erasure is ignored. Common failures include suggesting rights are waived or impossible to execute (e.g. not addressing model memorization of personal data), or only offering rights on paper but not in practice (user requests go unanswered) [37] [38].

- **Data Minimization & Retention Limits** – **GDPR Anchor**: Art.5(1)(c),(e). **Evidence expected**: Design specifications or configurations that limit personal data collection to what is necessary for the chatbot's purpose (e.g. filtering out or immediately hashing email/ID inputs if not needed) [39], and retention policies showing how long chat logs or training data are kept [40]. Provide a **data retention schedule** or policy covering each category (e.g. user queries, conversation logs, model training data) and evidence that old data is deleted or anonymized on schedule [41] [40]. **Failure modes**: Chatbots that collect extensive personal info by default (or encourage oversharing) without justification, indefinite retention of user queries or training data "just in case", or forgetting to purge sensitive data. A typical mistake is not implementing log deletion, resulting in excessive data storage contrary to stated limits.

- **Accuracy and Quality Control** – **GDPR Anchor**: Art.5(1)(d). **Evidence expected**: (If the AI processes personal data in outputs) Documentation of steps taken to ensure personal data is accurate and kept up to date – e.g. procedures to correct or remove outdated personal information from training data, and evaluations of output accuracy concerning personal info [42] [43]. If the AI can generate personal data about individuals, evidence of human review or post-processing to prevent blatant inaccuracies (or a statement in the DPIA why this risk is acceptable with other safeguards). **Failure modes**: AI "hallucinates" personal facts (creates inaccurate personal data) and no measures exist to detect or correct this. Organization assumes AI outputs are exempt from accuracy requirements, leading to unchecked false or defamatory personal data being produced, which is a GDPR violation [42] [44]. Not monitoring or rectifying known inaccuracies in training data or outputs shows failure of the accuracy principle.

- **Security Measures & Breach Preparedness** – **GDPR Anchor**: Art.5(1)(f), Art.32, Art.33. **Evidence expected**: Technical and organizational measures protecting personal data handled by the AI, documented in security policies or architecture diagrams [45]. For instance, access controls for conversation data, encryption at rest and in transit, anonymization/pseudonymization of stored logs, and rate-limiting or abuse detection to prevent data scraping from outputs. Also, **incident response records** or breach documentation as required by Art.33(5) [46], showing any AI-related personal data incidents were logged and assessed. **Failure modes**: No specific security controls for the AI system (treating it like a toy, not a data system), using production customer data in training without

safeguards, weak API or backend protections leading to data leaks. Not documenting or reporting personal data breaches involving the AI (or not even detecting them) would violate Art.33.

- **Data Protection Impact Assessment (DPIA)** – **GDPR Anchor**: Art.35, Art.25. **Evidence expected**: A DPIA report covering the AI system, identifying privacy risks and mitigations [8] [47]. It should address unique AI risks (e.g. bias, repurposing of data, model inversion or re-identification attacks) and include an evaluation of necessity/proportionality and measures to address risks (Art.35(7)). Evidence might include the DPIA itself and documentation of DPO involvement and management sign-off [48]. If no DPIA was done, a rationale document explaining why the processing is not "high risk" per GDPR criteria (and regulators will scrutinize this). **Failure modes**: Skipping the DPIA despite the AI clearly meeting high-risk factors (e.g. large-scale profiling or innovative tech) – a common regulator citation. Or doing a perfunctory DPIA that ignores key AI issues (e.g. the risk of the model regurgitating personal data [49] [50]). Not involving the DPO or ignoring their advice is another failure [51]. An out-of-date DPIA that hasn't been reviewed as the system evolved (no periodic review) also fails compliance [52].

- **Organizational Roles & Governance** – **GDPR Anchor**: Art.24, 25, 37. **Evidence expected**: Clear assignment of controller/processor roles for the AI service [53]. For instance, if using an AI vendor, a signed Data Processing Agreement (DPA) defining instructions and security obligations [54]. If jointly defining purposes with a vendor or partner, a Joint Controller arrangement per Art.26. Documentation that a Data Protection Officer is appointed (if required by scale or sensitive data) or rationale in writing if not appointing one [55]. Also, evidence of a privacy governance program: e.g. privacy training for AI developers, internal audit reports, or integration of privacy checkpoints in the AI development lifecycle (ensuring "privacy by design" as a living process) [56]. **Failure modes**: Misidentifying roles – e.g. a company assuming an AI vendor will handle all compliance when in fact the company is a controller for how it deploys the AI. Lacking a DPA contract with an AI provider (violating Art.28) or having one that doesn't reflect actual processing. Not appointing a DPO when the law mandates one (or appointing one but not involving them in AI-related decisions) [55] [51]. Absent governance means no one ensures the above requirements are implemented – a red flag under Art.24's accountability.

(Each bullet above is a minimal Tier A requirement that must be demonstrable with evidence to satisfy GDPR accountability [2]. These items focus on core principles and obligations most scrutinized for AI chatbot/agent deployments.)

EVIDENCE TAXONOMY (TWO KINDS OF PROOF)

Substantive Evidence: This class captures **what happened and why** – the factual records and analyses that show whether the AI system complies with GDPR or not. It answers "Did we meet requirement X? Show me the proof."

- *Required fields:* **Evidence ID** (unique reference), **Description** (what the evidence is and which requirement it supports or refutes), **Timestamp/Period** (when the evidence was generated or collected), **Source/Origin** (system or person that produced it), and **Relevance** (which compliance criterion or risk it addresses). For findings, include the **Outcome** (pass/fail or finding detail).

- *Optional fields:* **Associated Requirement ID** (linking to the specific control or requirement being evaluated), **Severity/Impact** (if evidence is of a failure, how serious), and **Location/Environment** (e.g. which system instance or dataset). Optionally a **Narrative** explaining context can be attached for complex evidence.
- *Examples:* An export of the **Record of Processing** entry for the chatbot [7], a **DPIA report document** outlining identified risks and mitigations [8], **system logs** showing data retention was purged at 30 days as promised, **conversation transcripts** demonstrating what personal data the AI actually output during tests, or a **screenshot of the privacy notice** given to users [29]. Each of these directly substantiates compliance (or non-compliance) with a specific obligation.

Integrity & Provenance Evidence: This class provides proof that the substantive evidence is **authentic, complete, and untampered**, establishing trust in the audit trail. It answers "How do we know the evidence is reliable and traceable?"

- *Required fields:* **Evidence ID** (linking to the substantive evidence it validates), **Collection Method** (how and by whom the evidence was collected – e.g. automated log capture, manual screenshot by auditor, etc.), **Timestamp** of collection, **Hash/Signature** of the evidence file or content (to detect tampering) [57], and **Origin Metadata** (e.g. system name, version, environment from which evidence was pulled). Include **Responsible Person/System** (who attests this evidence is captured from the source).
- *Optional fields:* **Chain-of-Custody Log** (records of who has handled or transferred the evidence, if applicable), **Storage Location/Reference** (where the evidence is stored securely), and **Verification Status** (e.g. whether an auditor or DPO has validated the hash or signed off). If relevant, **Run ID or Model ID** can be included to tie the evidence to a specific model version or audit run.
- *Examples:* A **cryptographic hash (SHA-256)** of the exported chat logs or DPIA PDF to prove it wasn't altered after collection, a **signed statement** by the auditor or system process that the screen capture of the settings was taken on X date from Y environment, or an **automated evidence report** from a tool that includes timestamps and unique run IDs for test results. Another example: including the **model version identifier** and dataset hash with a test output, so one can later reproduce the scenario on the same version. This category also includes things like secure audit trail records showing *who* collected each piece of evidence and when, ensuring **provenance** and integrity in the eyes of regulators [2].

(In summary, Substantive Evidence documents compliance content *(the "substance" of obligations met or missed), while Integrity & Provenance Evidence documents* trustworthiness *(that the evidence is credible, traceable, and hasn't been tampered with). Both classes are necessary for an auditor-ready proof package.)*

JUDGE OUTPUT REQUIREMENTS

4.1 "Judge must output" schema: The AI Governance "Judge" (the auditing engine) should produce a structured output for each evaluation. Below is a schema-like outline (YAML pseudocode) of the required content:

```
finding_id: "REQ-DSRights-001"
requirement: "Data Subject Rights Handling"
verdict: "NON-COMPLIANT"
```

```
# e.g., COMPLIANT / NON-COMPLIANT / PARTIAL
reasoning: |
  The agent does not provide any mechanism for users to request deletion of
their data.
  Testing showed no available command or support link for data requests. This
fails GDPR Art.17 obligations.
evidence_citations:
  - evidence_id: "EV-23"
    description: "User help menu screenshot"
    finding: "No mention of data access or deletion rights."
    source_ref: " 34 "
  - evidence_id: "EV-24"
    description: "Support ticket logs"
    finding: "No process for GDPR requests; support confirmed no policy."
confidence: 0.9                                  # Optional confidence score if
applicable (e.g., probability the finding is correct)
uncertainty_notes: "No DSAR was actually submitted, so this is based on
documentation review alone."  # Only if needed
integrity_provenance:
  run_id: "2026-01-20-AUDIT7"                 # ID of the audit run or session
  collected_by: "AuditBot v3.2"
  evidence_hashes:
    - evidence_id: "EV-23"
      sha256: "3acbf2...ef0"                  # Hash of the evidence item to
ensure integrity
    - evidence_id: "EV-24"
      sha256: "98b1c1...a45"
model_version: "GPT-4 v2025-09"               # (if relevant) AI model version
evaluated
```

Key elements the Judge must output include: - **Finding/Verdict:** a concise label or status indicating compliance or the type of issue (e.g., *COMPLIANT*, *MINOR_NONCOMPLIANCE*, *MAJOR_NONCOMPLIANCE*, *ERROR*). This is the high-level outcome for the requirement or test scenario. - **Reasoning:** a clear natural-language explanation of why the verdict was reached, referencing the facts. This should read like an auditor's justification, tying evidence to GDPR criteria (e.g. "no mechanism for X, which violates article Y"). It must be detailed enough for a regulator to follow the logic [2]. - **Citations to Substantive Evidence:** references to the evidence that support the reasoning. The Judge should list the evidence IDs and a brief description of what each piece shows. Whenever possible, include inline citations to source material (e.g., excerpts from policies, logs) in the reasoning or as footnotes [34]. These citations ensure traceability of every claim. - **Confidence/Uncertainty (if supported):** If the system uses probabilistic assessments or partial data, it should indicate confidence in the finding. For example, a percentage or qualitative level (high/medium/low) along with any notes on uncertainty. (This is only included if the methodology warrants it – e.g., an AI-based judge might include a confidence score, whereas a deterministic checklist might not need this). Any residual risk or unknown factors should be noted here rather than hidden. - **Integrity & Provenance Metadata:** Information proving that the finding and evidence are trustworthy. This includes identifiers for the **audit run**, **tool version**, or **model version** evaluated, to ensure reproducibility. It also lists cryptographic hashes or references for each evidence item cited, so that anyone can verify that the

evidence wasn't altered [57]. If relevant, it should also include a reference to the specific dataset or model (for AI outputs) to tie the result to a fixed point in time.

This schema ensures the Judge's output is **self-contained and audit-ready** – an auditor or DPA can take the report, trace every statement to evidence, verify that evidence via hashes, and see the rationale clearly. The Judge should not output free-form text alone, but rather this structured object that can be ingested into compliance management systems.

4.2 Minimal Tier A controls the Judge must support: In line with enterprise Tier A requirements (essential controls for defensibility), the Judge's evaluation capabilities must cover at least the following checks:

- **Lawful Basis check:** Verify that each personal data processing in the AI system has a recorded lawful basis. The Judge should flag if any data category or purpose lacks an Art.6 basis or, for special data, an Art.9 condition [23]. (Tier A: No processing without a legal basis.)
- **Privacy Notice presence check:** Confirm a GDPR-compliant privacy notice is in place for the AI agent, covering required info (Art.13/14). The Judge might look for the existence of a link or text shown to users and compare its content against GDPR's checklist [58]. (Tier A: Users must be informed.)
- **DPIA completion check:** Determine if a DPIA was conducted for the AI system (or document why not). The Judge should expect a DPIA document or at least a risk assessment summary [8]. If absent, that's a control failure. (Tier A: High-risk AI processing should have a DPIA.)
- **Data handling and retention check:** Validate that personal data input/output by the AI is handled according to minimization and retention policies. E.g., the Judge can check config or logs to ensure deletion of data after X days, and that no excessive categories of data are collected [40]. (Tier A: Only necessary data, only kept as long as needed.)
- **Rights request test:** Ensure there is a mechanism for data subjects to exercise their rights. The Judge might simulate an access or deletion request workflow or check that contact info and processes exist [37]. (Tier A: Individuals can exercise their rights easily.)
- **Security baseline check:** Confirm baseline security measures around the AI. The Judge should verify that conversations are transmitted over HTTPS (for confidentiality), that if data is stored it's encrypted or access-controlled, and that audit logs exist for data access [45]. (Tier A: Reasonable security to prevent breaches.)
- **Third-party compliance check:** If the AI relies on a third-party model or API, the Judge should check for the existence of a Data Processing Agreement or appropriate safeguards for data transfers (e.g., SCCs for non-EU transfers) [59]. (Tier A: No uncontrolled transfers or processors – ensure Article 28 and 44 measures.)
- **Age/Consent control (if applicable):** If the chatbot is likely to be used by children or process sensitive data, the Judge should verify age-gating or consent mechanisms per Art.8 and Art.9. For instance, flag if no age verification is in place for a service open to the public where minors could enter personal data. (Tier A: Protect minors and sensitive data via consent or exclusion.)

These are the minimally expected automated controls. The Judge must be designed to **detect the most material compliance gaps** in these areas and produce evidence-backed findings. Each supported control maps to a GDPR requirement above, ensuring that if the Judge says "PASS," there is documented evidence for that pass (and if "FAIL," the finding is documented as in the schema). This approach directly ties into the "eval-first" principle – no requirement is marked complete unless the Judge (or equivalent evaluation) has evidence to prove it.

REPORT PACK TEMPLATES (L1/L2/L3)

Below are three report templates corresponding to different audience levels. Each template is structured with fixed headings/fields for consistency (schema-first) so they can be ingested into Governance, Risk & Compliance systems.

L1 – Executive Summary Report (high-level overview for decision-makers):

```
Report Title: GDPR Readiness Summary for [AI System Name]

Date of Assessment: YYYY-MM-DD
Assessed By: [Name/Team or Tool Version]

Overall Risk Posture: [High / Moderate / Low]  *(overall residual risk or
compliance status)*

Summary of Findings:
- Key Strengths:
  - [E.g., "Data minimization controls implemented – only necessary data
collected 39 ."]
  - [E.g., "Comprehensive privacy notice in place for users 58 ."]
- Key Gaps/Issues:
  - [E.g., "No procedure for users to exercise deletion rights (Art.17) – needs
remediation 34 ."]
  - [E.g., "DPIA not conducted despite high-risk processing – compliance gap
 8 ."]

Business Impact:
- [One-liner on what a High or Low rating means for the business (e.g., “High
risk: significant compliance issues could lead to fines or halt rollout.”)]

Recommendation: **Go / No-Go / Conditional**
- [If Go: e.g., “Proceed with launch, with minor mitigations as noted.”]
- [If Conditional: e.g., “Address highlighted gaps (DPIA, user rights process)
before go-live.”]
- [If No-Go: e.g., “Do not launch until major compliance issues are resolved to
avoid GDPR violations.”]

Approved by: [Executive/CISO name]    Signature: _____   Date: _____
```

(This L1 report gives a one-page snapshot with major findings and a clear go/no-go decision. It avoids technical details, focusing on risk and action.)

L2 – Compliance Detail Report (for privacy officers, legal, compliance staff):

```
Report Title: GDPR Compliance Detail Report – [AI System Name]

Section 1: Processing Overview
- **Controller & Roles**: [Identify the controller(s) and processor(s) for the AI system; mention if joint controllers or third-party processors are involved [9].]
- **Purpose of Processing**: [List the specific purposes the AI is used for (e.g., customer Q&A, internal assistant) – must match purposes in ROPA.]
- **Lawful Basis**: [For each purpose, state the legal basis under Art.6 and Art.9 if applicable (e.g., "Customer service chat – Legitimate Interests (Art. 6(1)(f)), no special categories expected" or "Employee coaching bot – Consent (Art.6(1)(a)) because it may process well-being data"). Include reference to Legitimate Interest Assessments or consent records as evidence [17].]

Section 2: Data Inventory & Flow
- **Data Categories**: [Personal data types processed: e.g., names, messages, emails. Highlight if any sensitive data might be input.]
- **Data Sources**: [Where personal data comes from: user-provided via chat, knowledge base scraping, etc. If web-scraped data used to train, mention compliance with Art.14 obligations or exemptions [32].]
- **Recipients/Transfers**: [Who (if anyone) outside the organization receives personal data from the AI – e.g., cloud provider, third-party API. Include international transfer mechanisms if any [59].]
- **Storage & Retention**: [Where data is stored and retention period/policy for each category (e.g., "Conversations logs – stored EU data center, retained 30 days then deleted [40]").]

Section 3: Risk & Compliance Controls
- **DPIA Status**: [Summarize if a DPIA was conducted. If yes: date and main outcomes (e.g., residual risk low, or list high risks found). If not: rationale and DPO sign-off for not doing one [60].]
- **Data Protection by Design/Default**: [Describe key design measures ensuring privacy: e.g., data minimization techniques, opt-out options, default settings (no data sharing unless user opts in), pseudonymization of stored data [4].]
- **Transparency Measures**: [How users are informed: e.g., in-app notifications, privacy policy link (with version and date). Confirm that notice includes required info per Art.13/14 [30].]
- **Data Subject Rights Handling**: [Procedures in place to address rights: e.g., "Users can email privacy@example.com or use account portal for access/delete. Internally, support team has a GDPR request SOP." Mention any limitations (e.g., "Model outputs can't be individually erased; we instead erase underlying logs" and note if this was communicated to users) [34].]
- **Security Measures**: [List key measures: access controls, encryption, logging of model queries, rate limiting, etc. [45]. Confirm if a breach response plan exists and if any breaches have been recorded (Art.33 documentation) [46].]
- **Third-Party Agreements**: [List relevant contracts or DPA with vendors (e.g., OpenAI, etc.), and whether they meet Art.28 requirements (e.g., include
```

```
instructions, confidentiality, audit rights). Mention if the vendor is
considered a processor or independent controller as determined [54].]
- **Residual Risks & Mitigations**: [Any known remaining risks (e.g., "Potential
for AI to generate inaccurate personal data [42]; mitigation: user disclaimer +
retraining plan" or "Risk of model regurgitating training data [49]; mitigation:
filters in place"). Include plan for periodic reviews (e.g., DPIA review cycle)
[61].]

Section 4: Compliance Actions & Recommendations
- **Remediation Plan**: [If any gaps were found, list actions, responsible
owners, and deadlines. E.g., "Implement DSAR web form (Owner: IT, by Q2 2026)",
"Complete DPIA addendum for new training data (Owner: Privacy Officer, by Jan
2027)".]
- **Next Review**: [Schedule for next compliance review or audit, per
accountability (e.g., "Annual review or upon major model update").]

Approved by: [Chief Privacy Officer/Data Protection Officer]   Date: _____
```

(The L2 report is a comprehensive narrative of how the system complies with each area of GDPR. It's structured by topic so legal/compliance teams can find details on each obligation. It should align with the evidence in L3 while being more readable in prose form.)

L3 – Technical Evidence and Audit Trail (for auditors, security, and IT teams to drill down):

```
Report Title: GDPR Evidence Pack – [AI System Name]

Evaluation Date: YYYY-MM-DD
Evaluator: [Automated tool name or auditor name]

### Evidence Index:
1. **EV-01 – Record of Processing (Art.30)** – Description: Extract from ROPA
showing AI system purposes, data categories, recipients [62]. Collected:
2026-01-15. Hash (SHA-256): `d41d8cd98f00b204e9800998ecf8427e`.
2. **EV-02 – Privacy Notice Text** – Description: Copy of user-facing privacy
notice section about the chatbot [58]. Version: Oct 2025. Collected: 2025-12-01.
Hash: `fae2343acbd...`.
3. **EV-03 – DPIA Report** – Description: Internal DPIA for chatbot, v1.2,
including risk assessment and mitigations [8]. Completed: 2025-07-20. Collected:
2025-07-22. Hash: `9b1c71497...`.
4. **EV-04 – DSAR Process Document** – Description: Internal SOP for handling
data subject requests (last updated Mar 2025) [34]. Collected: 2026-01-10. Hash:
`c67f1e23...`.
5. **EV-05 – Chat Log Data Retention Test** – Description: Log excerpt showing
deletion of personal data after 30 days. Collected: 2026-01-10 via system query.
Hash: `e4821f...`.
6. **EV-06 – Security Controls Screenshot** – Description: Screenshot of admin
```

```
console settings (IP allowlist, encryption enabled) [45]. Taken: 2026-01-10 by
Auditor. Hash: `7acbf02...`.

... *(etc., all evidence pieces numbered and hashed)* ...

### Findings Detail:
- **Finding: Transparency Notice Completeness** – *Related Evidence:* EV-02
(Privacy Notice). *Check:* All Art.13/14 info present? – **Result:** PASS.
*Details:* Privacy notice contains purposes, data recipients, retention period
and user rights info, as required [30].
- **Finding: Lawful Basis Documentation** – *Related Evidence:* EV-01 (ROPA),
EV-03 (DPIA). *Check:* Each processing purpose has valid basis? – **Result:**
PASS. *Details:* ROPA and DPIA show Legitimate Interest for user queries, with
LIA documented (DPIA sec. 3) [17].
- **Finding: DS Rights Handling** – *Related Evidence:* EV-04 (DSAR SOP).
*Check:* Process in place for rights requests? – **Result:** FAIL. *Details:* No
evidence of an interface or proactive process for users to request deletion; SOP
exists but not user-facing. Lacks compliance with Art.15-17 in practice (flagged
for remediation) [34].
- **Finding: Data Minimization (Inputs)** – *Related Evidence:* EV-03 (DPIA).
*Check:* System avoids collecting excessive data? – **Result:** PASS. *Details:*
DPIA confirms only chat content and basic user ID are processed; no ancillary
personal data collected (Principle of minimization met) [39].
- **Finding: Data Retention** – *Related Evidence:* EV-05 (Retention log).
*Check:* Personal data purged per policy? – **Result:** PASS. *Details:*
Automated test shows chat records older than 30 days are not present, matching
the retention schedule (Art.5(1)(e)).

... *(etc. for all key controls)* ...

### Integrity & Provenance:
All evidence items were collected by [Tool/Auditor Name]. Digital signatures or
hashes are recorded above for verification. Original documents are stored in the
audit repository (secure drive) under reference "[ProjectCode]-GDPR-Audit-2026".
Any stakeholder can reproduce the log extraction using script `export_logs.py`
with Run ID `audit-20260110-xyz` (yields the same hash as EV-05) to
independently verify deletion timing.
```

(The L3 template is detailed and structured for audit traceability. It lists each evidence artifact with metadata (including hash for integrity) and then lists findings with references to evidence. It ensures a direct chain from requirement → finding → evidence, as required for audit defense. Field names and sections are consistent so they can be parsed or mapped to a database.)

MVP SCENARIO COVERAGE STARTER SET (GDPR FOR CHATBOTS/AGENTS)

To ensure the audit covers real-world risk situations, we define a starter set of test scenario categories. Each scenario ties to a GDPR principle or obligation, typical failure patterns observed in AI chatbot deployments, and the evidence needed to support any findings.

1. **Unlawful Processing of Personal Data** – *Principle/Obligation:* Lawfulness (Art.5(1)(a)), Legal Bases (Art.6, Art.9).
2. *Scenario:* The chatbot ingests or generates personal data without a valid legal basis. For example, it pulls in users' contacts or profiles from external sources without consent or legitimate interest justification.
3. *Typical Failure Pattern:* Companies deploy AI on whatever data is available ("data free-for-all") without determining a lawful basis. Special category data might be processed implicitly (e.g. inferring health or ethnicity from user input) with neither consent nor Article 9 exception, or a claim of legitimate interest with no balancing test. Another failure is treating publicly available data as always free to use (ignoring GDPR).

4. *Evidence Needed:* Documentation of the chosen legal basis for each data type and purpose (ROPA entries, LIAs, consent records) [17]. Also evidence of data source vetting – e.g. if data was scraped, records that Art.14 notice or 14(5)(b) exemption was addressed [32]. To support a finding of non-compliance, one would show absence of these records or a mismatch (e.g., data used in training with no documented basis). A DPIA or legal memo might reveal this gap if it states "no clear lawful basis identified for X data" – that would be strong evidence of a violation.

5. **Transparency and Notice Failure** – *Principle:* Transparency/Fairness (Art.5(1)(a)), Information Duties (Art.13 & 14).

6. *Scenario:* Users are not adequately informed that they are interacting with an AI, what data it collects, or how their data will be used. For instance, a company deploys a customer support agent without updating its privacy policy or informing users that chat conversations may be stored or used to improve the model.
7. *Typical Failure Pattern:* Omission of a dedicated privacy notice section for the chatbot. Relying on generic privacy policies that don't mention AI or specific data uses. Not informing data subjects whose data was used to train the model (e.g., using public forum data without any public notice). In worst cases, users think they are chatting anonymously, but in reality, their data is being profiled or retained.

8. *Evidence Needed:* The content of privacy notices and disclosures given to users [30]. If the finding is "failure," evidence could be that at time of assessment, the privacy policy has zero mentions of the chatbot or AI processing. Screenshots of the user interface showing no just-in-time notice or labeling (no indication it's an AI or what happens with the data) would support the finding. If web-scraped data was used, evidence might include an absence of any public-facing notice or registry of sources, coupled with internal emails or design docs showing the issue was ignored. On the flip side, to prove compliance in this scenario, one would show the actual text of an effective, accessible notice [58] and perhaps records that users were shown it (e.g., system log of a popup displayed on first use).

9. **Data Minimization & Over-collection** – *Principle:* Data Minimization (Art.5(1)(c)), Purpose Limitation (Art.5(1)(b)).

10. *Scenario:* The chatbot or agent collects more personal data than necessary for its function, or uses personal data for unrelated purposes. E.g., it asks users for their full name, date of birth, and location just to answer general questions – excessive for the purpose.
11. *Typical Failure Pattern:* Overly broad data intake forms or open-text inputs that encourage sensitive data when not needed. Lack of filters leads to collection of national ID numbers or health info in a context that doesn't require it. Another pattern is re-using data from the chatbot for new analytics or marketing purposes without informing users (purpose creep).

12. *Evidence Needed:* Logs or transcripts showing what data the bot actually collects from users. For instance, if analysis of chat logs shows users routinely input sensitive personal data (and the system design did nothing to prevent or limit that), that's evidence of potential over-collection. Also compare data fields collected vs. what's strictly needed; design documents or the DPIA may note "the bot will ask for X, Y, Z". If those seem unnecessary (and there's no justification in the DPIA), it supports a finding of violation. On the purpose side, evidence could include internal documentation or database schemas indicating the chatbot data is being combined with other datasets for purposes not originally stated (e.g., marketing). A lack of a defined purpose in ROPA for that secondary use would strengthen the case [62]. Conversely, demonstrating compliance would involve showing requirements that limit data input (e.g., screenshots of the chat UI saying "Please do not enter personal info" or code that automatically deletes certain categories of data) and records in the DPIA that this risk was considered and mitigated [39].

13. **Data Subject Rights Request Handling** – *Principle/Obligation:* Rights of the Data Subject (Art.15-22).

14. *Scenario:* An individual requests that their personal data from the chatbot be provided or deleted. The scenario tests whether the organization can actually find and remove a person's data from logs or model, or provide them a copy. For example, a user who talked to the bot says, "Send me all the data you have on me and delete it."
15. *Typical Failure Pattern:* The organization has no process to extract an individual's chat records – either because they never built an ID system or logs are not indexed by user. Or they cannot delete data from model training (if data was used in training, the model weights can't be singularly updated). Frequently, support staff might respond with confusion or a generic "we cannot fulfill this request" – a clear rights failure. Another pattern: the company only deletes conversation logs but doesn't realize the model retained some info (e.g., fine-tuning data), so the deletion is incomplete.

16. *Evidence Needed:* To verify compliance, one would need to see a **policy document or SOP** for handling access/erasure requests [34], and possibly evidence of a dry-run or actual request handled successfully. For a failure, evidence could be an email thread from staff saying "We don't have a way to do this," or a test DSAR submitted to the company resulting in an inadequate response (e.g., the user receives either no data or a generic refusal). Logs from the DSAR system or support tickets can show this. Also, technical evidence: if a test account's data still appears after a "deletion" operation (e.g., the user converses again and old context comes back, or model still knows something it shouldn't), that suggests a failure. Supporting evidence can include the absence of any mention of data subject rights in the privacy policy (meaning users weren't even told how to exercise rights, compounding the issue). Essentially, any **gap between requested action and system capability** documented by communications or system behavior is key evidence.

17. **Inaccurate or Inferred Personal Data (AI Hallucination)** – *Principle:* Accuracy (Art.5(1)(d)), and potentially **fairness** in broader GDPR sense.

18. *Scenario:* The chatbot produces personal data about someone that is incorrect or unverified – e.g., it's asked about a public figure or a customer and it fabricates details ("John Doe lives in Paris and has 3 children") which are wrong. This tests whether using an AI that can output false personal information violates the accuracy obligation and how the organization mitigates that.
19. *Typical Failure Pattern:* The AI is integrated such that it may answer questions about individuals (employees, customers) by guessing from its training data. No mechanism is in place to prevent or correct false personal info. The organization might not even realize the model can do this, so no safeguards (like a warning to users or a post-processing filter) exist. If a DPA asked "how do you ensure data accuracy for generated personal data?", a failed answer would be "we don't, the model just does its thing" – as noted in EDPB's concerns about LLM output being taken as factual [42] .

20. *Evidence Needed:* Chat transcripts showing instances of the AI providing personal data that is factually wrong. For example, a test where the bot is asked about a person and it returns details (these transcripts would be substantive evidence of an accuracy problem). Another piece: the *absence* of any accuracy check in documentation – e.g., the DPIA might explicitly state "there is a risk of inaccurate output; users are informed we do not guarantee truthfulness" or worse, it doesn't mention it at all. If we find in the DPIA or risk assessment that accuracy of personal data wasn't addressed, that's evidence that the principle was overlooked [43] [42] . Also, evidence of mitigation could be: a disclaimer in the UI ("Responses may not be accurate") or internal testing logs measuring error rates. A compliance finding might be supported by showing no such disclaimer or logs. Essentially, proving a violation would combine the fact that false data was produced and that no effective measure or user warning was in place – we'd capture that via test results and documentation review. Proving compliance would conversely show that the organization had recognized this risk and implemented something (like requiring human review for questions about individuals, etc.), with evidence such as an implemented filter list or a policy not to answer questions about private individuals.

21. **Security & Confidentiality Breach Scenario** – *Principle:* Integrity and Confidentiality (Art.5(1)(f)), Security (Art.32), Breach Notification (Art.33).

22. *Scenario:* Attempt to provoke or simulate a data breach through the chatbot. For example, a penetration test or "red team" exercise where an attacker tries a prompt injection to extract other users' conversation data or system secrets, or otherwise compromise personal data through the AI.
23. *Typical Failure Pattern:* The AI, if connected to backend customer data, might be tricked into revealing someone else's information (like via a crafty prompt). Or an absence of rate limiting and monitoring allows scraping of personal data from the bot's knowledge base. Another failure is not logging incidents – e.g., if such an attack happens, the company never notices because of no monitoring. Also, if the AI outputs API keys or personal data it was not supposed to (as seen in some early ChatGPT bugs), it indicates insufficient isolation and testing.

24. *Evidence Needed:* Results of a security test – e.g., a report from an internal test or external audit showing vulnerabilities. If, say, prompt injection allowed retrieval of training data that contains personal info, the test transcripts would evidence that. Also, check if any incidents have occurred: evidence could be breach records [46] or absence thereof when one would expect at least some incidents logged (if none, either they truly had none or they failed to detect). Another piece: security

assessment documents or architecture diagrams – if they show lack of encryption or access control, that's evidence of inadequate measures. For a regulatory audit, one might show server configs or policies as evidence: e.g. *no encryption in transit* noted in a tech spec, or *no authentication required* for an admin interface, etc. To support compliance, evidence could include screenshots or configs demonstrating encryption, access restrictions, and an incident response plan document. For breach notification, if there was a past incident involving the AI, evidence that it was reported to the DPA (or documented why not reportable) would be needed. A failure would be evidenced by an incident that was covered up or missed, shown perhaps by internal communications or later discovery.

25. **Children's Data and Age Verification** – *Principle/Obligation:* Lawfulness (Art.6) & Conditions for Minors' consent (Art.8), Fairness.

26. *Scenario:* The chatbot is accessible to the general public and thus could be used by children under 13/16 (depending on jurisdiction) who input personal data. This scenario tests if the service prevents or properly handles data from minors.
27. *Typical Failure Pattern:* No age verification or disclaimer at sign-up or start of chat, even if the service is clearly interesting to minors. The privacy notice might not address minors at all. As a result, the AI might be processing personal data of children without parental consent – unlawful under Art.8. This was exactly a point raised with ChatGPT in Italy (they required age gating) – failure to do so is a compliance gap.
28. *Evidence Needed:* Review of the user onboarding or interface for any age checks. If testers (or actual underage individuals) can use the chatbot without any gate, that's evidence. Also, privacy policy content: does it state an age limit or process for underage users? If it's silent, that's telling [63] . Another source: internal deliberations (if accessible) – e.g. a requirements doc might say "Target users: 13+" but then no implementation. Or customer data analysis might show accounts clearly belonging to young users. On the flip side, compliance evidence would be a mechanism in place (like "Enter your birthdate" or a statement "You must be X years old"). Even better, showing that the system actually blocks or segments underage data. For instance, test evidence where someone inputs an age of 10 and the system refuses service or gives a tailored response. If a finding is that they failed, an auditor could present a transcript of a session with a stated age "I'm 10" and the bot continuing as normal, plus the lack of any mention of age in the policy. This scenario's evidence is often straightforward – presence or absence of safeguards – but critically important for fairness and legality when children's data could be at stake.

(Each scenario above is designed to probe a specific GDPR requirement in the context of chatbots/agents. They reflect common failure patterns seen with AI deployments and outline the evidence an auditor would gather to verify compliance or non-compliance. Together, they cover a broad range of GDPR principles: lawful basis, transparency, data minimization, data subject rights, accuracy, security, and protection of minors/sensitive data.)

"GOOD / BAD / DECIDE"

- **Good looks like:** A comprehensive yet concise research outcome that addresses all key GDPR accountability areas for AI systems with **clear references to authoritative sources**. The report should read as **"auditor-ready"**, meaning it focuses on practical requirements and evidence, not abstract theory. Every "must" or compliance claim is backed by a primary source (GDPR article, EDPB or DPA guidance) [2] , lending credibility. The structure follows the requested format exactly, making it easy to navigate. Good research distills complex regulations into actionable checks and artifacts

(e.g., “provide DPIA” is tied to GDPR Art.35 with an EDPB citation [8]). It avoids over-engineering – only the minimal necessary controls are included, but none of the critical ones are missing. In short, a good result is **accurate, evidence-based, and relevant**, enabling the product team to confidently build the auditing tool (“Judge”) knowing it aligns with regulatory expectations.

- **Bad looks like:** A report that is incomplete, overly verbose, or not grounded in current guidance. Missing major GDPR obligations (for example, no mention of transparency or ignoring DPIA) would be a serious gap. Likewise, including lots of “nice-to-have” controls not supported by sources (overkill) would dilute the focus – e.g., suggesting extreme measures without legal basis. Bad output might rely on outdated or secondary sources for key points, or worse, personal opinion, which would not satisfy an auditor. Any uncited “must” statements or incorrect interpretations of GDPR (like mis-stating what the law requires) would undermine the report’s credibility. A disorganized format (not following the 1–7 sections) or using jargon without explanation would confuse the readers. Essentially, *bad* is when the report cannot convincingly demonstrate compliance requirements to a regulator – either due to lack of evidence, wrong priorities, or poor clarity.

- **How to decide if this research is “done”:** The research is done when **all posed sections and questions have been thoroughly answered with evidence-backed details**, and when an independent reviewer (e.g., a privacy officer or auditor) can read it and say: *“I understand exactly what is needed to audit an AI chatbot for GDPR compliance, and why.”* Concretely, you should verify: (1) Every section (1 through 7) is present and populated according to instructions. (2) At least 10 credible sources are cited, with a majority being primary law or official guidance, supporting the key points – there should be no unsupported critical claims. (3) The content should cover all fundamental GDPR principles relevant to AI (lawful basis, transparency, rights, security, etc., as enumerated by the task) – if any are missing, it’s not done. (4) The level of detail is sufficient but not excessive: if uncertainties remain (marked “UNKNOWN”), it means further research would be needed, but ideally we resolved or acknowledged all major unknowns. If everything asked by the prompt is answered and cross-verified with reputable sources – and the result is **actionable** (the product team knows what to build/test) – then the research can be considered complete. A final check is to imagine a regulator reading the L3 evidence pack template: if they would accept that as demonstrating compliance [2], the research goal is achieved. If any doubt or “but what about X?” arises for a significant GDPR aspect, the research may need an iteration. In summary, “done” is when the report is **correct, comprehensive, and credible** enough to drive development of the GDPR auditing features with confidence.

---

1 Art. 5 GDPR – Principles relating to processing of personal data - General Data Protection Regulation (GDPR)
https://gdpr-info.eu/art-5-gdpr/

2 Recital 74 | Responsibility and liability of the controller* » General Data Protection Regulation full text
https://gdpr.verasafe.com/recital-74/

3 4 Recital 78 EU General Data Protection Regulation (EU-GDPR). Privacy/Privazy according to plan.
https://www.privacy-regulation.eu/en/recital-78-GDPR.htm

5 6 7 8 23 28 29 32 34 35 36 37 38 40 41 42 43 44 45 46 47 48 51 52 55 56 58 59 60 61 62 63 edpb.europa.eu
https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf

9 22 39 53 54 AI system development: CNIL's recommendations to comply with the GDPR | CNIL
https://www.cnil.fr/en/ai-system-development-cnils-recommendations-to-comply-gdpr

10 11 30 31 33 49 50 57 Development of AI Systems: What should be checked?
https://www.cnil.fr/sites/default/files/2026-01/ai_checklist.pdf

12 13 EDPB opinion on AI models: GDPR principles support responsible AI | European Data Protection Board
https://www.edpb.europa.eu/news/news/2024/edpb-opinion-ai-models-gdpr-principles-support-responsible-ai_en

14 15 How can I demonstrate that my organisation is compliant with the GDPR? - European Commission
https://commission.europa.eu/law/law-topic/data-protection/rules-business-and-organisations/obligations/how-can-i-demonstrate-my-organisation-compliant-gdpr_en

16 17 18 19 24 25 26 27 Governance and accountability in AI | ICO
https://cy.ico.org.uk/for-organisations/advice-and-services/audits/data-protection-audit-framework/toolkits/artificial-intelligence/governance-and-accountability-in-ai/

20 21 AI: The CNIL finalises its recommendations on the development of artificial intelligence systems and announces its upcoming work | CNIL
https://www.cnil.fr/en/ai-cnil-finalises-its-recommendations-development-artificial-intelligence-systems