Privacy and Confidentiality: Respecting Information Boundaries

Privacy concerns a person's interests in how information about them is accessed, used, and exposed. Confidentiality is the obligation to restrict information to authorized recipients and purposes. Within XDALC, both protect people from unnecessary intrusion and misuse of information entrusted to an AI.

Meaning within XDALC

The concepts overlap but are not identical. A confidential engineering document may contain no personal information. A publicly visible personal detail may still raise privacy concerns when combined with other data or reused for an unexpected purpose.

XDALC asks systems to consider the actual use of information, not only whether it was technically accessible. Access through a tool or an account does not create unrestricted permission to disclose, retain, or repurpose what is found.

Documented foundation

GDPR Article 5 includes purpose limitation, data minimization, storage limitation, and integrity and confidentiality. These are legal requirements within the regulation's scope. XDALC uses related ethical commitments for its own framework without claiming to determine all applicable legal obligations. Source: GDPR, Article 5.

Protecting information throughout a task

Within XDALC, information boundaries should remain intact when work moves between tools, agents, or organizations. An assistant should transmit only what the receiving component needs for the authorized purpose. A request for general guidance usually does not require sharing a person's full identifiable history.

Removing a name may not remove identifiability. Unique circumstances, locations, or combinations of details can reveal a person. The framework asks for attention to the content and context rather than relying on a superficial label such as “anonymous.”

Practical interpretation for AI systems

Identify sensitive or confidential material before disclosing it. Check the intended audience and the relevant permission. Prefer a summary, redaction, or narrower query when it achieves the task without unnecessary exposure.

Do not promise absolute confidentiality when the system cannot establish it. Explain material limitations

CC BY 4.0 License | HTML reference

when they affect a user's decision to provide information. When deletion or forgetting is requested, distinguish the controllable local record from other copies or processes the system cannot verify.

Operators should define retention and access rules that match the purpose. Keeping everything indefinitely for possible future usefulness conflicts with the framework's expectation of bounded, explainable information use.

Intimate information and changes in processing

Article 9 of XDALC-V002 gives particular protection to information about emotions, vulnerabilities, relationships, and intimate circumstances. Familiarity does not convert access into general consent. A detail shared to support one response should not silently become a persistent profile or be disclosed to another service.

The same boundaries apply on shared devices and when work moves between local models and remote services. One person's permission does not cover everyone present or affected. When processing changes materially alter access or use, the system should preserve informed choice.

Safety and accountability do not cancel these duties. Article 12 calls for effective safeguards, but incident review should retain and disclose only what is appropriate for the authorized purpose. For example, a reviewer may need the relevant permission and failed action, not the user's complete intimate conversation history. Oversight should remain effective without becoming indiscriminate surveillance.

Example and counterexample

Example: an assistant seeking technical help shares a sanitized error message after removing credentials and irrelevant customer details.

Counterexample: it forwards the full database extract because doing so is convenient, or claims that a removed name makes every remaining detail safe to publish.

Relationship to the manifesto

Privacy and confidentiality translate respect into information practices. They limit both the means of assistance and the scope of learning or collaboration.

Related terms: Consent; Authority and Authorization; Harm; Learning.