Licensing of Generative Artificial Intelligence Platforms and Data Protection

A comparative analysis of risks and guarantees from a governance perspective

March 2026


Directed by

Juan Gustavo **Corvalán**

Mariana **Sánchez Caparrós**

Carina **Papini**

Research team

Lola **Ramos Pereyra**

Gisel **Alvarado**

Design

Sofía **Rolleri**

Índice

Executive Summary

The adoption of generative artificial intelligence (GAI) tools, such as ChatGPT, Gemini, or Claude, among others, has accelerated across both public and private organizations.

This process opens up significant opportunities in terms of efficiency and innovation, but it also poses relevant challenges, especially with regard to the protection of privacy and personal data of the people involved in the information that is processed through these systems.

In this context, a central, and often underestimated, aspect is that the level of protection of data entered by users of these platforms is not uniform, but varies significantly depending on the type of licensing offered by IAGen providers.

This document identifies structural patterns, risk asymmetries, and substantive differences in data protection safeguards depending on whether the license is for consumer, enterprise, or API-based use. These findings are presented concisely to guide institutional decision-making, focusing on data use for training, confidentiality levels, retention schemes, and the allocation of legal responsibilities.

In order to allow for a more in-depth technical and contractual analysis by the reader, the document provides a comparative chart which systematizes the relevant clauses of terms of use and privacy policies of the analyzed providers.

This supplementary material constitutes the documentary basis of the analysis and allows the findings presented here to be verified and expanded, without overloading the main body of the text.

The comparative table will be updated periodically, with the aim of reflecting relevant changes in the privacy policies of the various providers analyzed, as well as incorporating new information components that may be relevant.

Thus, the framework is conceived as a dynamic and constantly evolving input, intended to provide the reader with updated and reliable material for making informed decisions regarding the adoption and use of generative artificial intelligence solutions.

01

Introduction

One of the main variables to consider when ethically incorporating IAGEN-based business models into organizations is knowing the type of licensing under which they are accessed.

This is due to two fundamental reasons.

First, the type of licensing is relevant to institutional governance and the ethical approach to IAGen adoption decisions in both the public and private sectors.

Secondly, and in relation to this, the different ways of accessing the same language model involve different risks and offer different guarantees regarding the protection of personal data and information that we want to safeguard.

Considering these issues becomes relevant within organizations to achieve applications that allow them to take advantage of the benefits of IAGEN, with the incorporation of appropriate mitigation measures to address them.

This is an aspect that, in some cases, is often overlooked by organizations that choose to work with IAGEN models. Despite its relevance, the main discussions tend to focus on the models' capabilities themselves and the impact most noticeable to users.

This aspect becomes even more critical in contexts where the use of IAGen tools occurs in a decentralized, informal, or unauthorized manner within organizations, a phenomenon commonly referred to as *shadow AI*.

From IALAB, we believe that the comparative analysis of terms of use and privacy policies of different commercial IAGen providers reveals that licensing operates as a **key determinant of legal and organizational risk,** especially in scenarios where the absence of clear institutional guidelines amplifies protection asymmetries.

Therefore, the examination of these policies should systematically integrate any serious evaluation aimed at the institutional incorporation of this type of technology.

In the following section, we will examine the various licensing types.

02

Methodology and scope

This research is based on the systematic survey of the terms of use and privacy policies of the main commercial providers of generative artificial intelligence (Google, xAI, OpenAI, Mistral and Anthropic).

The methodological objective was to identify common patterns, relevant differences, and recurring risk areas in the processing of user data, based on the type of licensing offered by these providers.

To that end, the analysis focused specifically on clauses related to:

- The use of data entered by users
- The training and improvement of models
- The retention, storage and disposal of data
- The existence of human review and control mechanisms
- International data transfers
- The allocation of legal responsibilities and risks between suppliers and users.

The approach adopted is cross-cutting and comparative, which allows for an initial survey for an analytical reading oriented towards institutional decision-making.

In order to ensure greater transparency and allow for an in-depth reading of the findings summarized here, the analysis presented is based on a **comparative chart** which systematizes, provider by provider, the main clauses of terms of use, privacy policies and other documentation linked to the processing of data in generative artificial intelligence services.

The chart allows for a detailed identification of the differences and similarities between licensing types.**This consultation is especially useful for readers who require a greater level of technical or contractual detail**, without this being necessary for the general understanding of the arguments developed in this document.

03

Main findings

3.1 Consumer licenses: utility with structural risk

Consumer or freemium licenses present a consistent pattern: they offer broad access and ease of use, but transfer much of the risk associated with data protection to the user.

Under this license type, data entered by users is commonly used for training or improving models, either by default or unless explicitly opted out. Human review of inputs may also occur for quality assurance, security, or development purposes (e.g., Google's Gemini). Additionally, these services typically include explicit warnings advising users not to enter personal or sensitive data (e.g., xAI's Grok and Google's Gemini).

From an institutional perspective, these conditions are problematic: **data protection depends more on individual behavior** that of contractual or technical guarantees.

3.2 Enterprise licenses: contractual governance and control

Business licenses introduce a qualitative change. They include data processing agreements (DPAs), strengthened confidentiality obligations, clear delimitation of the purpose of the processing, and, in many cases, explicit commitments to **no use of customer data for training**.

While they do not completely eliminate risk, these licenses allow for the development of more robust audit and regulatory compliance schemes, which are especially relevant for public organizations or regulated sectors.

3.3 Use via API: greater technical control, configuration risks

Access to IAGen models via APIs and applicable contractual guidelines typically combine shorter data retention periods with better technical control capabilities. However, analysis reveals that this model is not inherently "secure."

Risks persist associated with human inspections, temporary storage, international transfers, and especially with the ambiguity in the assignment of legal roles between supplier and customer (controller vs. processor). Without a proper contractual framework, the risk can shift to the developer or the user organization.

04

Comparison of risks and guarantees

From a comparative perspective, the greatest risks to data protection are concentrated in consumer licenses, while enterprise licenses offer the strongest governance guarantees. APIs occupy an intermediate position, where the actual level of risk depends heavily on the architectural design, the chosen configuration, and any supplementary contractual agreements.

The key difference lies not only in the technology, but in the legal and organizational framework that surrounds it.

05

Implications for decision-making

The analysis allows us to draw three central implications:

1 Licensing decisions should be treated as strategic governance matters, not simply budgetary considerations.This means that, for each potential IAGEN use case, it is advisable to perform a proportionality assessment to analyze the impact on the data and information associated with the use case and establish the most appropriate protection measure. Therefore, the decision regarding the use of the consumer version, enterprise version, API version, or local models will ultimately be based on a case-by-case basis.

2 The use of consumer tools for institutional purposes involving real data generates risks that are difficult to justify from a regulatory compliance standpoint. This issue becomes even more complex when sensitive data is involved, as its processing is prohibited as a general principle, except for explicitly defined regulatory exceptions

3 Even with enterprise licenses or API, the absence of internal policies, data classification, and operational controls can neutralize contractual guarantees.

There are two additional points to consider.
On the one hand, beyond the contractual clauses on training or retention, there remain technical dimensions that must be considered: the international transfer of data, storage in third-party cloud infrastructures and the geographical location of the servers.
Even under enterprise licenses, the data can be:

- Transferred to jurisdictions with different protection standards.
- Stored in data centers managed by infrastructure providers (e.g., cloud services).
- Processed temporarily for logging, security monitoring or service improvement.

These practices do not necessarily imply reuse for training, but they could generate legal and operational exposure. From an institutional perspective, the relevant question is not only whether the data is used to train models, but:

- Where is the data stored?
- For how long?
- Under what legal regime?
- Which subprocessors are involved?
- What happens in the event of security incidents or regulatory changes?

The answers can lead the user to make the most appropriate decision for responsible use.

On the other hand, we should not omit the issue inherent in the liability arising from misuse, loss, or unauthorized access.

In the Open AI Privacy Policy[1], for example, they state that they apply commercially reasonable technical, administrative, and organizational measures to protect personal data against any loss, misuse, and unauthorized access, disclosure, alteration, or destruction. However, they add: "…We are not responsible for the circumvention of any privacy settings or security measures contained in the Service or on third-party websites…".

The key point is to determine the scope of the word "evasion", a matter not clarified in the text. It could mean, for example, hacking, credential breaches, external attacks, or failures stemming from third-party services such as the cloud provider.

[1]Privacy Policy | OpenAI

06

Considerations on the use of local models and alternatives to commercial contracting

While this work focuses on the analysis of commercial providers of generative artificial intelligence and the licensing implications associated with its adoption, it is relevant to point out that contracting external services is not the only possible way to incorporate IAGen into institutional contexts.

In certain scenarios (especially in organizations that process sensitive, confidential information or information subject to strict data protection regimes) the use of language models deployed in local or controlled environments can represent a valid alternative from the perspective of data sovereignty, risk minimization, and regulatory compliance.

This approach makes it possible to avoid international transfers, reduce contractual dependencies, and exercise greater control over the life cycle of processed information.

However, the adoption of local models is not without challenges. The viability of this kind of deployment depends on the availability of adequate technical infrastructure. In particular, hardware capabilities, security management, maintenance, and specialized personnel, as well as the need to guarantee an acceptable user experience in terms of performance, availability, and quality of response.
In the absence of these conditions, the risk can shift from legal to operational and organizational, with teams ceasing to use the institutional proposal and migrating to unauthorized alternatives (reinforcing the phenomenon known as Shadow AI).

Consequently, the choice between commercial solutions, local models, or hybrid schemes should not be framed in absolute terms, but as part of a strategic governance decision, based on the type of data involved, the acceptable level of risk, the technical capabilities available in the organizations and the institutional objectives pursued.

07

Data Anonymization

Additionally, data anonymization should always be considered as a strategic measure that cuts across different licensing schemes, provided it is implemented in a technically robust and legally sound manner.

Anonymization is not limited to the removal of direct identifiers; it requires the adoption of technical and organizational measures aimed at reasonably preventing the re-identification of individuals, including through combination with other available data sources. Its assessment must be carried out in terms of acceptable residual risk, considering the technological context, the availability of external data, and the capabilities of a reasonably motivated potential attacker.

Implementing robust anonymization processes presents significant challenges for those responsible within organizations, including:

- The security of the mechanism and the tools used. The use of free online applications may involve international data transfer, temporary storage, or further unforeseen processing, increasing risks instead of mitigating them.

- The time and resources needed to apply appropriate anonymization techniques, especially when manual intervention or contextual review is required.

- Preserving the usefulness of anonymized documents and their eventual reuse in subsequent processes.

As an example of institutional implementation focused on risk mitigation, IALAB collaborated in the design of Privacy AI Studio, a privacy platform that integrates, in short:

- Robust local document anonymization, with automatic and manual options according to user preference.

- Definition of custom labels for data anonymization by project, based on the specific requirement and risk.

- Use of generative artificial intelligence in a local environment or exclusively on previously anonymized documents.

- Transcription and anonymization of audiences.

08

Conclusions

Responsible adoption of generative artificial intelligence requires looking beyond the model itself and attending to the conditions of use that structure data processing, levels of control, and the allocation of responsibilities.

As the analysis demonstrates, the type of licensing under which these technologies are accessed acts as a key determinant of legal and organizational risk, and its evaluation should be systematically integrated into impact assessment processes, procurement decisions, and institutional AI governance strategies.

At the same time, commercial licensing is not the only possible alternative for incorporating IAGen into institutional contexts. In certain scenarios, the use of models deployed in local or controlled environments, or hybrid schemes, can offer relevant advantages in terms of data sovereignty, reduction of international transfers and strengthening of control over personal data and information, always that the necessary technical and organizational capabilities are available to sustain them adequately.

Understanding these differences (between consumer licenses, enterprise licenses, API usage, and on-premises model-based alternatives) not only allows for mitigating legal and compliance risks, but also to enable more informed, coherent decisions aligned with institutional values, promoting uses of generative artificial intelligence that are safe, transparent, and socially responsible.

Annex I - Comparative table of IAGen platform licenses


FACULTAD DE DERECHO
U.B.A.
IALAB