Whitepaper

The AI sovereignty imperative

From digital dependency to organizational control

eraneos

Contents

Executive summary

Over the past decade, major advances in AI have allowed for the automation of work that previously required human labor. AI systems can now drive cars, read books, write articles, talk, listen, program, and make scientific discoveries. They are even beginning to think. Organizations now use AI in its many forms in their core business practices to increase efficiency, reduce costs, and improve the quality of their products and services. We foresee that this trend will continue, and that organizations not using these new technologies will be at a competitive disadvantage.

While the reliance on foreign, primarily United States Cloud Service Providers (USCSPs), for core data and infrastructure is a known challenge, a new, equally critical dependency has emerged: Artificial Intelligence (AI) services.

The rapid progress of AI, particularly large language models (LLMs), is changing how organizations operate, making LLMs an important factor for competitiveness. However, most innovation and commercialization in this field is driven by a few foreign companies, which introduces profound sovereignty risks. When we outsource core thinking and decision-making to machines controlled by foreign entities, one key question arises: **How can we ensure that AI works in our organization's interests and not those of external parties? How can we access these crucial capabilities that will increasingly help us operate our business?**

In this era of AI outsourcing, digital sovereignty — the ability to control and govern one's digital infrastructure, data, and technologies — is a necessary countermeasure. In our paper, we define digital sovereignty using four key principles: Sovereignty of purpose, Inspection, Adaptation, and Provider choice.

The reliance on closed, proprietary AI-as-a-Service models also introduces five serious risks: Alignment risk (where the model's behavior may subtly bias decisions to serve foreign interests); Confidentiality & Privacy risks (as all input/output data is transferred and stored abroad, potentially violating compliance laws like GDPR); Availability risk (where critical operations can be paralyzed by service denial); Integrity risk (where providers can deploy a secretly handicapped or censored model); and Compliance risk (which is heightened by foreign surveillance legislation like the US CLOUD Act).

Although the technology has developed rapidly, the debate over AI sovereignty has significantly hindered its adoption in Europe. European organizations, particularly in the public sector, have been reluctant to embrace new technologies due to concerns about technological autonomy. As a result, AI adoption has stalled, putting Europe at a competitive disadvantage.

A core dilemma emerges as the most capable LLMs today pose the highest sovereignty risk as a result of their closed nature. To mitigate this, organizations must move away from a "trust and don't verify" approach by adopting a set of clear strategies. The most effective path to high sovereignty is self-hosted, open-weight AI, which allows organizations to run verified models on sovereign infrastructure, fully mitigating risks around availability, confidentiality, and integrity. Other key strategies include implementing AI-provider independent architectures to enable seamless switching and prioritizing EU-native AI service providers for regulatory alignment.

At Eraneos, we understand that regulations can only do so much. As such, we focus on the strategic imperative of achieving digital sovereignty for AI services, guiding corporates in getting the most from their AI solutions without compromising on control. In this paper, we explore the unique risks posed by foreign AI dependency and provide concrete strategies for organizations to mitigate these threats and move toward an independent and sovereign AI posture.

Understanding AI sovereignty

The majority of cutting-edge AI services, particularly advanced LLMs, are delivered by a small group of foreign-based, highly capitalized providers, predominantly from the United States (such as OpenAI, Google, xAI and Anthropic). These AI services, and your organizational data – both the input sent to the models and the output they generate – are only available with the consent and support of the provider. Ultimately, the third-party provider of the AI solution is in control. That's why it's important to understand what AI sovereignty means, its strategic importance, and its relevance in the broader context of digital sovereignty.

The strategic imperative of AI sovereignty

Do these third parties have your complete trust, especially, with AI systems increasingly driving strategy and operations? Moreover, organizations need to ask themselves a few important questions about these providers:

- **Will they reliably continue service** even in the face of political pressure? (Availability risk)
- **Do they keep data confidential** from third parties, model training and even from intelligence agencies? (Confidentiality & Privacy risks)
- **Can the AI model's behavior be manipulated** or censored without your knowledge? (Integrity & Alignment risks)
- **Can your organization comply with laws and regulations** that govern your data and decision-making? (Compliance risk)

Recent geopolitical tensions between the EU and the US have fundamentally altered the trust equation, not just for EU member states but for all EU organizations. This calls for a critical reappraisal of the dependency relationship our society and its organizations have with these dominant, foreign AI service providers.

In this whitepaper, we have outlined a definition of digital sovereignty in the context of AI services. We have also identified opportunities for achieving greater control, higher agility, better resilience, and long-term strategic advantage through sovereign AI strategies.

Defining digital sovereignty in the age of AI

Digital sovereignty is the ability of an organization to control and govern its own digital infrastructure, data, and technologies. In the context of AI, this means ensuring that the complex AI solutions that increasingly drive an organization's strategy and operations align with the organization's unique interests, values, and jurisdictional requirements.

For AI services, digital sovereignty manifests through four key principles:

1. **Sovereignty of purpose**
The organization must be able to freely choose how and for what purpose to use the AI system without fear of coercion or interference from external parties (e.g. the AI provider or its governing state). This ensures the AI systems ultimately serve the organization's goals.

2. **Sovereignty of inspection**
The organization must have the ability to understand, inspect, and audit the AI system. This is crucial for verifying alignment, preventing the introduction of bias, and ensuring the model is not serving adverse hidden interests.

3. **Sovereignty of adaptation**
The organization must be able to adapt the AI systems to its evolving needs and unique context. Dependence on a closed, proprietary model severely restricts this ability, making the organization inflexible.

4. **Sovereignty of provider**
An organization must be able to freely choose and switch between AI service providers, preventing dependence and lock-in to a single, powerful vendor. This mitigates vendor pricing power, availability risk and allows for an exit strategy when the risk assessment of a provider changes.

Assessing sovereignty: readiness levels and risks

Risks of AI-driven digital dependency

The reliance on foreign AI services introduces specific and serious risks that stem from the inherent nature of AI models, their immense training costs, and the resulting concentration of power among a few US-based providers. We identify five core sovereignty risks around AI technologies:

1. Alignment risk
This is arguably the most insidious risk. LLMs and other advanced AI systems are trained and refined to exhibit or inhibit certain behaviours. A foreign AI provider can, without your knowledge, subtly guide your decision-making, withhold information on certain topics, or introduce biases that align with the provider's priorities or those of its government. If we outsource our thinking and decision-making, we must ensure the machine's "interests" are aligned with our own. The opaque, "black-box" nature of closed proprietary models makes verifying true alignment impossible.

2. Confidentiality & Privacy risk (data leakage)
When an organization uses an LLM-as-a-Service model, all data sent to the model (input) and all data generated by it (output) is received and stored by the foreign provider. This sensitive data could potentially be used - without your explicit consent - for further AI training, research, reselling, or even be offered to foreign intelligence services and data brokers, as enabled by legislation like the US CLOUD Act. Even promises of non-use are unverifiable.

The use of AI services, particularly those handling sensitive customer or employee data, must comply with strict regulations like the GDPR in the EU. Relying on foreign providers whose data practices are subject to non-EU jurisdictions (e.g. the US Foreign Surveillance Act) makes it extremely difficult, if not impossible, to guarantee compliance and protect citizens' privacy rights, exposing the organization to significant legal and financial risk.

3. Availability risk (service denial)
As AI services become a core part of organizational workflows, the dependency on their continuous operation grows. A foreign AI service provider, just like a cloud service provider, can deny service for political, sanction-related, or accidental reasons. This could instantly paralyze critical business processes, increasing dependency and decreasing sovereignty.

4. Integrity risk (model manipulation)
Foreign AI providers maintain control over the underlying model's weights and configuration. They can decide - without your knowledge or ability to notice - to deploy a handicapped or censored model based on your identity, workplace, or geography. This integrity breach means the AI's output could be subtly altered or incomplete in ways that actively work against your interests, leading to poor operational or strategic outcomes. Given that the provider operates outside your organization's jurisdiction, detecting or diagnosing such manipulation without their active help is impossible.

Digital sovereignty readiness levels for AI

To move from digital dependency to full digital sovereignty when making use of AI services, an organization must first accurately determine its current position on the dependency-sovereignty spectrum. This assessment is crucial for defining targeted sovereignty ambitions and actively working toward them.

We divide the continuum of AI sovereignty into four readiness levels:

Level	Description	**AI Sovereignty Implications**
0	**Dependent** 	The organization uses closed, proprietary AI-as-a-Service models from foreign providers (e.g. US-based). Sovereignty risks are unknown or unanalyzed. The reliance is based purely on trust, with no measures in place to mitigate Alignment, Confidentiality, Availability, Integrity, or Privacy risks.
1	**Prepared** 	The organization is dependent on foreign AI service providers but has identified and analyzed its key sovereignty risks. Measures are taken to reduce current sovereignty risks and the organisation is prepared to move towards level 2 (independent) in the future with minimal disruption. Data confidentiality, integrity, and alignment risks are known but for now remain based on unverified trust.
2	**Independent** 	The organization is only partially dependent on foreign AI providers. Sovereignty risks are actively mitigated (e.g. using EU Native AI Service Providers or highly-protected, end-to-end encrypted workflows). Business continuity is ensured. Critical and privacy-sensitive data sent to the AI is protected, and the integrity of the process is verifiable to some extent.
3	**Sovereign** 	The organization is almost independent of foreign AI service providers for its continuous operation (e.g. using self-hosted Open-Weight AI on sovereign infrastructure). Sovereignty risks are continually monitored and actively managed. The organization has control over the AI model weights, inference process, hardware, software, and internal expertise, effectively mitigating all five core risks.

In the following sections, we will describe specific strategies for increasing your AI digital sovereignty readiness level.

Sovereign AI service strategies

The use of AI technologies is thus quickly becoming a must-have for organizations around the world. The majority of innovation and commercialization of AI technologies and services is driven by foreign companies from the US and China. This introduces a very real sovereignty risk: if we outsource our thinking and decision making to machines, and these machines are under control of foreign companies and governments, how can we ensure the machines work in our interests and not theirs?

The five core sovereignty risks with regard to AI technologies are:

- **Availability:** What is the risk of the AI service provider ceasing service?
- **Confidentiality:** Is the data sent to and generated by the AI model confidential?
- **Integrity:** Can the data processed by the AI model, the model itself or the AI-applications be changed without your knowledge?
- **Privacy:** Can you be in compliance with privacy regulations when using AI?
- **Alignment:** Will the AI model work in your interests or in someone else's?

Recent advancements have established LLMs as the forefront of the technological disruption driving these risks. Because the most capable models are typically offered as a managed service, an organization's most sensitive data and decision-making processes are continuously exposed to these risks. Specifically, the nature of how LLMs are created and deployed directly translates into heightened Alignment Risk, Confidentiality Risk, and Integrity Risk, making a focused analysis of their characteristics essential for developing a robust sovereignty strategy.

AI sovereignty for Large Language Models

Before LLM risks can be truly guarded against, organizations must fully understand exactly what an LLM is.

An LLM is a sophisticated AI system engineered to process, comprehend, and generate complex human-like text by drawing upon massive volumes of language data. These models are trained on diverse sources - spanning books, articles, and vast web content - enabling them to master linguistic patterns, grammar, factual knowledge, and intricate reasoning abilities. The result is the capacity to understand user prompts and deliver highly coherent, contextually appropriate responses. Significantly, modern LLMs are rapidly expanding their utility through multi-modal capabilities, allowing them to interpret and generate documents, images, audio, video, and program code.

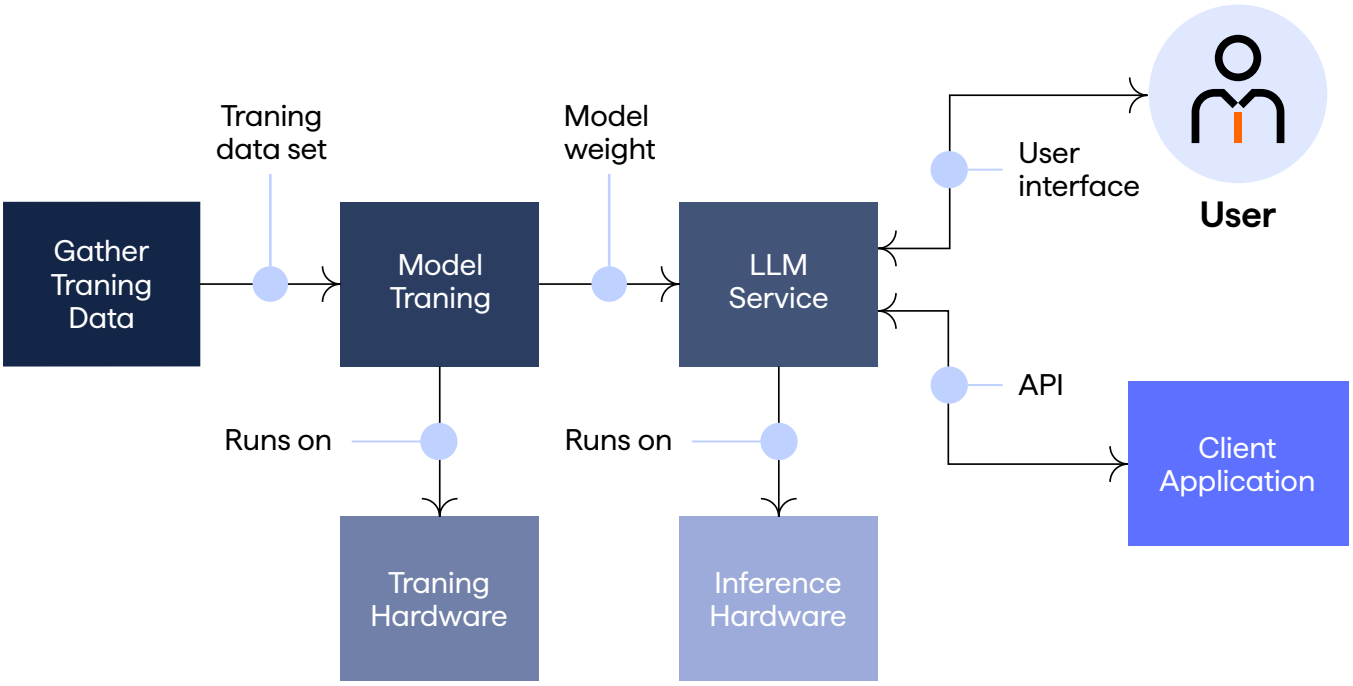
At their core, LLMs operate via deep learning, a machine learning approach that leverages simulations of neural networks. A neural network processes information by passing the input through multiple layers of interconnected simulated neurons. Each connection possesses a numerical value known as a weight. During the training phase, these weights are automatically calibrated based on the provided input data and the desired output. When the network is subsequently run (a process known as inference), these learned weights are loaded into memory, and new input is processed through the structure to produce the final output.

How Large Language Models are made

The lifecycle of an LLM is a multi-stage engineering endeavor. It commences with the critical step of gathering a large, high-quality dataset - the designated training data. This data then informs the subsequent phases: designing a robust model architecture and creating a sophisticated model training system.

The core of the process involves training the model on this system using accumulated training data, followed by refining the model through various post-training techniques (e.g. alignment and safety protocols).

The utility of the LLM in production requires two further critical steps: running the LLM (a process known as inference) and efficiently serving the model to end-users or client applications via a dedicated LLM service interface. This entire pipeline necessitates significant capital and specialized expertise.



The capital and compute requirements necessary to train and maintain a state-of-the-art LLM are enormous, rendering this undertaking unfeasible for all but the most heavily capitalized global organizations. Consequently, the resources to develop and commercialize the most advanced models are highly concentrated among a small number of providers, predominantly based in the United States.

This market dynamic dictates the availability model: the best-performing LLMs are closely guarded proprietary models. Access is almost exclusively offered through a consumption-based LLM-as-a-Service model, exemplified by platforms such as ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google) and Grok (xAI). This reliance on closed, external services inherently amplifies digital dependency and the associated sovereignty risks.

AI LLM classification: A spectrum of openness

To effectively assess and mitigate sovereignty risks, we must classify Large Language Models (LLMs) based on their degree of openness. This classification directly correlates with the control an organization can exert over the technology.

- **Open-data:** The entire dataset used to train the model is freely and publicly available. This allows for full verification of provenance and reduced bias risk, representing the highest level of sovereignty.
- **Open-training:** The code and methodology used for training the model are publicly accessible. This permits deeper inspection and allows for full reproducibility of the training process.
- **Open-weights:** The crucial model weights (the output of the training process) are freely available for download and use. This is the necessary condition for self-hosting and full integrity validation, mitigating risk significantly.

Conversely, closed-weight models keep these weights private, accessible only via a managed service API or proprietary interface. This model-as-a-service approach creates the highest dependency.

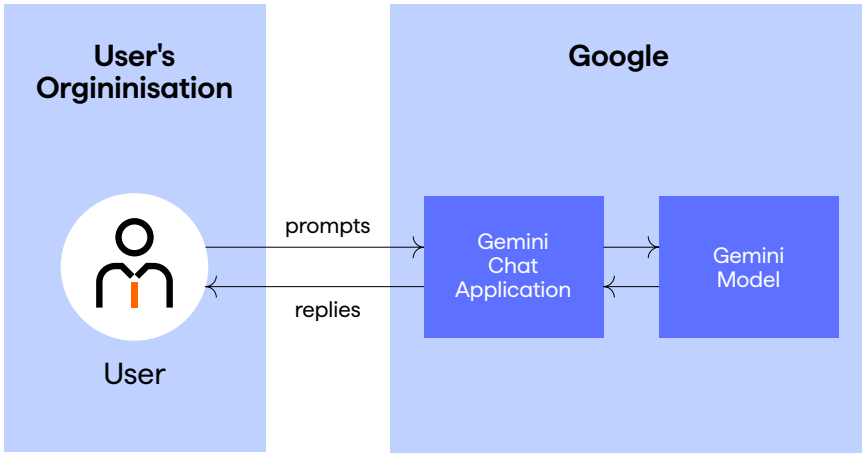
The current market reality is that the most capable, state-of-the-art LLMs – the very models organizations seek for competitive advantage – are predominantly offered as closed-weight models. This proprietary, managed service approach inherently dictates a fundamental loss of control over the model's operation, behaviour, and the data flowing through it. It is precisely this structural dependency on foreign providers for the core weights and inference process that introduces and amplifies the following critical sovereignty risks.

AI applications & deployment

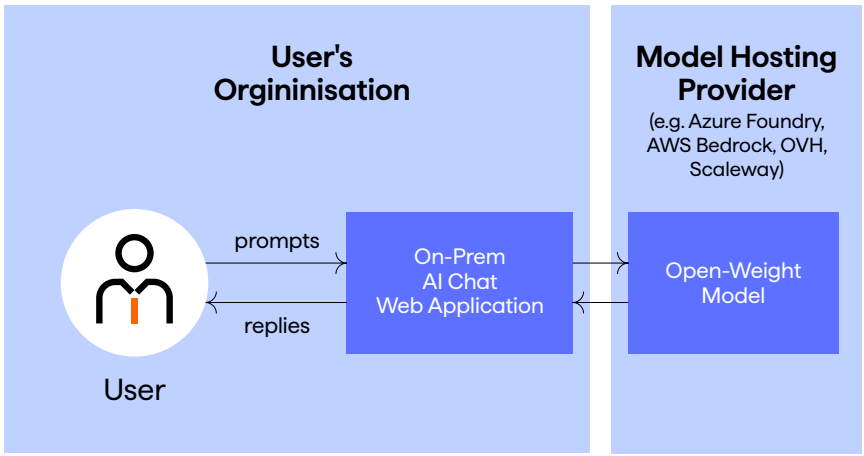
After an AI model is trained it is deployed on servers equipped with specialized hardware capable of executing the model. Communications with the AI-model work via an API.

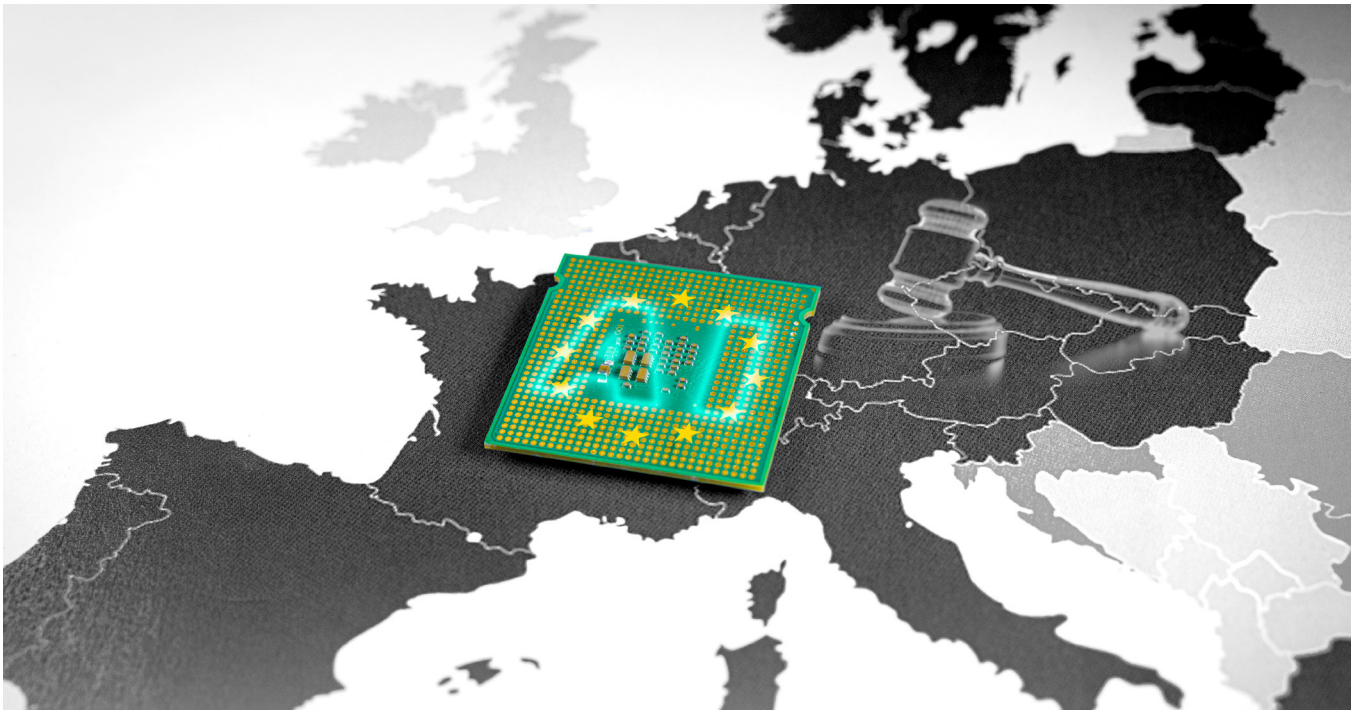
This AI model is then used by an AI Application like a user-facing chat interface, an autonomous agent or AI enabled software. The AI-application uses the AI-model via an API.

Many user facing AI applications like ChatGPT, Gemini and Claude are a combination of an hosted application and a hosted model trained by the same provider. For example the Gemini chat interface is developed and hosted by Google which uses Gemini Large Language models that are both trained and hosted by Google. In such a scenario the user's organization is highly dependent on one party – Google.



But it is important to realize that the application, the model deployment and the model training can be separated. The application could be created and hosted by one party, while the model is hosted by another party which was trained by yet another party. For example one can use an on-prem hosted proprietary or open-source chat application like to use an open-weight model trained by an AI Lab hosted on an model hosting provider. In this scenario the organization is free to choose another model hosting provider, deploy the open-weight model and switch over with minimal user – and thus business continuity – disruption.





Crucial choices in AI sovereignty

Four crucial choices impact AI sovereignty and the preparedness to move towards higher AI sovereignty in the future:

1. The choice of AI model
2. The choice of AI model hosting provider
3. The choice of AI application
4. The choice of AI application hosting provider

The choice of AI model

Open-weight AI models enable free choice of model hosting provider, which enables a safe exit route, provider redundancy, reduces provider lock-in and pricing power and reduces the availability risk. Also the ability to control the model version, which prevents surreptitious model updates reducing the integrity and aligment risk. The ability to fine-tune the model allows adjusting the model's performance to the needs of your organisation.

Open-data & open-training AI models reduce the dependency on the AI model developer. Because the data and the training method is open it's possible for other organisations to retrain the model. Additionally it's possible to validate the training method and data to have confidence in the model's alignment.

The jurisdiction of the model developer's company and the controlling entities is also be a concern. Model developers in other jurisdictions (e.g. the US) might be influenced by the interests of their company, parent company or nation-state, and these interests might not be align with your own. The developer of the AI model could – by force, threat or choice – train the model to exhibit certain behaviour. For example many AI-models trained by Chinese labs refuse to discuss politically sensitive topics. It is also possible to train models to alert authorities when certain behaviour has been detected.

Choosing a model trained by a model developer that is in an allied jurisdiction e.g. within the EU reduces the risks associated with foreign actors influencing your AI.

The choice of AI model hosting provider

The model hosting location locality is a factor in data locality. In- and output data is sent to/from the model's hosting location.

The model hosting providers entity jurisdiction impacts which parties and governments can influence the model hosting provider e.g. to shutdown services to specific companies, countries or organizations. Is the controlling entity based in country, in the EU, in the US or elsewhere.

The choice of AI application

Proprietary AI applications e.g. chat platforms or AI-enabled applications are closed source. This means you need to trust, but cannot verify that the software does not impact data integrity or confidentiality. Additionally you are dependent on the software supplier for the continuity of use. Suppliers under a foreign jurisdiction could be forced by their governments with sanctions (or the threat thereof) to discontinue delivery, service and maintenance. They can also be secretly forced to share data or install backdoors by foreign intelligence services.

Free and open-source (FOSS) AI applications (e.g. chat webinterfaces, agents or AI-enabled applications) offer the freedom to deploy, inspect and change the application. This reduces the confidentiality risk by allowing you to verify that the software does not leak confidential information or performs user tracking or monitoring. It also reduces the availability risk as the software is free and open source, so there is no supplier that can deny or restrict your usage. The integrity risk is also reduced because you can verify the software does not surreptitiously alters data going to or from the AI models. Additionally you can change the software to your liking yourself or hire any other third party to do this for you.

Choosing Free and Open Source AI applications increases the sovereignty of your AI stack.

The choice of AI application hosting provider

When an AI-Application is hosted outside your organisation all data going to and from the application can be intercepted, monitored and gathered by the hosting provider or parties with privileged access to the hosting provider.

Choosing a trustworthy hosting provider for AI applications is thus important confidentiality or integrity breaches. Choosing an AI application that can be hosted at multiple (or any) hosting provider provides you with an exit strategy and freedom of provider choice.

A maximalist AI sovereignty strategy would thus be:

1. **AI model:** an open-weight, open-training, open-data AI models from an AI model developer in a trusted jurisdiction
2. **AI model hosting provider:** on-prem or with a trusted AI model hoster within a trusted jurisdiction with more than one provider
3. **AI application:** a free and open source AI applications that is hosted in a trusted environment
4. **AI application hosting provider:** on-prem or with a trusted hosting party

Improving AI sovereignty

Much of AI use today is based on closed-weight, closed-training, closed-data AI models trained by a foreign AI-lab, via proprietary AI applications, both hosted on foreign cloud service providers, maximizing AI sovereignty risks.

But this is a choice-by-default, not a practical necessity.

During your AI transformation there are real, practical, no-regret choices you can make now to prevent the AI sovereignty risks from impacting you in the future:

1. Require freedom to choose an AI provider

When selecting, building or sourcing AI (enabled) applications require that the application can function with more than one AI model and more than one AI model provider. This makes you independent of the AI model developer and the AI model provider, allowing you a simple exit strategy if the risk assessment changes. Prefer on-prem over in country over in EU over US over elsewhere when it comes to the data locality and the jurisdiction of the controlling entity.

2. Require freedom to choose an AI applications hosting provider

When selecting, building or sourcing AI applications, require that the application can be hosted on any hosting provider of your choice. This enables you to shift hosting providers if the current hosting provider becomes unavailable, unwilling or too risky to use. Prefer on-prem over in country over in EU over US over elsewhere when it comes to the data locality and the jurisdiction of the controlling entity.

3. Prefer applications that are free and open-source

When selecting AI applications, prefer free and open-source software to give you the ability to audit the software and change or extend it to your needs. This also makes you developer independent; you can contract anyone to improve or maintain the software should the need arise.

4. Prefer open AI models

Require that AI application can work with models that are open-weight. This allows you to switch AI model hosting provider without having to switch the model. Prefer models that are open-trained and open-data. Adhering to these four principles when choosing your AI applications and models strikes a pragmatic balance: you are not independent, but you are not locked into dependence. This at the very least allows you to change providers, models and applications at will when the future risks turn into acute crisis.

Implementing a sovereign AI strategy today

Achieving full AI sovereignty is currently not feasible for most organizations, but improving the AI sovereignty is possible for all organizations. The following three practical measures can change your AI sovereignty level from dependent, through prepared to independent. Each measure can be implemented independently, or together.

Measure 1: AI provider agnostic chat

When using AI webchat like ChatGPT, Claude or Gemini, the AI application (ChatGPT), the models (GPT) and the model hosting provider (OpenAI) are all provided by one supplier. This increases the sovereignty risks: all data, availability and model behaviour are all controlled by a single supplier.

These risks can be reduced by using a 3rd party AI chat application that is supplier agnostic. This way you control the interface and can choose freely between multiple AI models and AI hosting providers. Additionally, this measure allows you to control the data, the integrations, the cost, and budgeting. A supplier-agnostic AI chat application can be hosted on a hosting provider of your choosing, further increasing your freedom of choice and reducing your dependency on a single party.

Measure 2: AI provider agnostic agents

Agents like Claude Code, Codex and Gemini CLI are tightly coupled to one model provider and their proprietary models. Selecting a provider agnostic agent allows you to freely choose both model providers and models.

Measure 3: AI provider agnostic AI gateways

A more general solution to ensure provider independence and control over data is to implement an AI gateway. A software service that relays AI applications to AI models. This can be deployed either as a SaaS or on a hosting provider of your own choice. An AI gateway serves as a proxy between your AI applications (like chat interfaces and agents) and model hosting providers. This ensures free choice of model provider and gives you control. You can for example ensure that certain AI applications use only certain (compliant) models and model providers, get insight and control over costs and seamlessly switch providers if necessary.

End-to-end services that deliver impact

Combining deep industry expertise with technological excellence, Eraneos delivers measurable business impact. A pragmatic approach ensures sustainable adoption over experimental implementation, while proven methodologies and front-runner status in Generative AI turn data challenges into strategic advantages – with full sovereignty and compliance.

Partner with Eraneos today to implement the AI sovereignty measures that fit your organization to ensure continuity, confidentiality and integrity in the AI age.

Author:

Mathijs de Meijer
Senior Consultant – Data & AI
Eraneos

mathijs.de.meijer@eraneos.com

Get in touch

Claudia Schulze
Partner – Data & AI
Germany
claudia.schulze@eraneos.com

Dave Kiwi
Practice Lead – Data & AI
Netherlands
dave.kiwi@eraneos.com

Katharina Fulterer
Partner – Data & AI
Switzerland
katharina.fulterer@eraneos.com

Antonio Rodriguez
Senior Manager – Data & AI
Spain
antonio.rodriguez@eraneos.com