

# An Open-Source Sovereign Cloud for an Open Europe: The Case for a Federated, AI-Enabled, and Multilingual Digital Infrastructure

*by Dinis Cruz and ChatGPT Deep Research, 2025/02/05*

## Executive Summary

The European Union (EU) stands at a critical juncture in defining its digital sovereignty. As reliance on U.S.-based cloud providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud deepens, the need for an independent, secure, and resilient cloud infrastructure has become an urgent priority. The ability to control and govern its own digital infrastructure is not just a technological issue but a matter of economic and political sovereignty. This white paper outlines a bold yet pragmatic vision for an **EU Sovereign Cloud Infrastructure**, designed to be fully open-source, seamlessly compatible with existing cloud services, and deployable at multiple levels—EU-wide, national, and even municipal.

A core principle of this initiative is **transparency and interoperability**. The proposed sovereign cloud must be **100% open-source** to ensure verifiability, security, and long-term sustainability, while simultaneously maintaining **full compatibility with major cloud APIs**. This approach enables organizations to migrate to the EU cloud without requiring major code rewrites or infrastructural changes, removing a significant barrier to adoption. Technologies such as **LocalStack**, which already emulates AWS services for local testing, demonstrate the feasibility of this approach—allowing cloud workloads to run seamlessly in different environments without the application ever realizing the difference. By adopting a similar model, the EU can achieve a **drop-in replacement** for existing cloud services, ensuring digital autonomy without disrupting ongoing business and governmental operations.

Beyond mere independence, this sovereign cloud strategy also recognizes the **economic and social benefits of decentralization**. Unlike monolithic cloud infrastructures controlled by a handful of corporations, the EU's sovereign cloud must be **federated**, allowing countries, regions, and cities to deploy and control their own cloud environments while remaining interoperable with the broader European network. This federated model fosters

local expertise, creates jobs, and strengthens the supply chain for cloud technologies across the continent. By distributing cloud resources more evenly, it also enhances resilience against outages, cyber threats, and geopolitical disruptions.

However, **digital sovereignty is not just about infrastructure—it is also about language and culture**. In the era of artificial intelligence (AI) and Large Language Models (LLMs), the dominance of English as the “default programming language” threatens to leave European languages at a disadvantage. AI models trained primarily on English datasets risk reinforcing linguistic and cultural biases, making technology less accessible and effective for non-English speakers. **This is why Portuguese, as one of the world’s most spoken languages, must be a central focus of Europe’s AI strategy.**

Portuguese, spoken by over **265 million people across four continents**, is not only a major European language but also an important global one. Yet, AI systems today struggle to process and generate high-quality Portuguese text at the same level as English. This discrepancy is not just a technical limitation—it is a **barrier to equal access**. If AI is to be the interface of the future, then ensuring linguistic diversity in AI systems is as crucial as ensuring sovereignty in cloud computing. This white paper proposes a strategic initiative to establish **Portuguese as a first-class programming and interaction language in AI**, ensuring that Portuguese speakers can engage with AI on equal footing with English speakers.

By treating **natural language as the new programming language**, AI has the potential to revolutionize how humans interact with technology. If well-developed AI models can “code” and perform complex tasks based on prompts in English, they should be able to do the same in Portuguese and other European languages. However, this will not happen by default—it requires deliberate investment in **Portuguese-language AI datasets, model training, and policy mandates** to enforce multilingual AI capabilities. Governments, businesses, and research institutions must collaborate to develop **native AI models** that understand and generate Portuguese as fluently as they do English, enabling new applications in education, governance, business, and innovation.

The alignment of **cloud sovereignty and AI linguistic diversity** presents Europe with a unique opportunity. A federated, open-source EU cloud infrastructure—coupled with AI that natively supports European languages—ensures not only **technological independence** but also **linguistic and cultural sovereignty**. By prioritizing **seamless API compatibility**, the transition to this sovereign infrastructure will be frictionless, enabling businesses and governments to move without disruption. And by treating Portuguese (and other European languages) as core programming and interaction languages, the EU ensures that its AI ecosystem serves all its citizens equitably, preventing an English-first AI future that disadvantages non-native speakers.

This vision is not just an aspiration—it is a necessity. The EU must act decisively, with **policy incentives, funding for open-source initiatives, and industry engagement** to make sovereign cloud and multilingual AI a reality. By embracing **Wardley Mapping as a strategic tool**, this white paper outlines an evolutionary roadmap for achieving these objectives, ensuring that Europe does not simply react to global technological trends but actively shapes them.

The path forward is clear: **a cloud infrastructure that is open, federated, and seamlessly compatible, and an AI ecosystem that is inclusive, multilingual, and equitable**. This will safeguard Europe's digital sovereignty, empower its diverse linguistic communities, and set a global example of a technology strategy that balances independence, accessibility, and innovation. Now is the time to take action—to build a future where European citizens and businesses are not just consumers of foreign technology but architects of their own digital destiny.

---

## Key concepts

- **EU Digital Sovereignty:** The European Union (EU) aims to develop a sovereign cloud infrastructure to ensure data security, economic resilience, and technological independence ([Sovereign Cloud Stack](#)). This effort responds to concerns about foreign control over critical data and services, and seeks to uphold European regulatory standards (like GDPR) without compromise.
- **Open-Source & API Compatibility:** This paper proposes a fully open-source, transparent cloud stack that maintains near 100% compatibility with existing cloud APIs (e.g., AWS, Azure) to enable seamless migration and usability. An open-source approach ensures transparency and autonomy, while API compatibility avoids the costly rewrites and integrations otherwise needed when moving off U.S.-based clouds ([EU Sovereign Cloud Initiative Drives Single-Source Solutions | Business Wire](#)) ([Sovereign Cloud Stack](#)).
- **Linguistic Inclusivity in AI:** Portuguese, along with other European languages, must be prioritized as a programming and interaction medium in AI development. Proactive support for these languages in AI will ensure linguistic equity and prevent over-reliance on English-centric AI systems. Portuguese is a globally significant language (over 265 million speakers ([World Portuguese Language Day | UNESCO](#))), and treating it as a first-class language in AI will help bridge current gaps where AI models perform worse in non-English languages ([ChatGPT Is Cutting Non-English Languages Out of the AI Revolution | WIRED](#)).

- **Strategic Evolution (Wardley Mapping):** Wardley Mapping principles are leveraged to outline the evolutionary path of EU cloud infrastructure and its strategic positioning. Wardley Mapping is a strategic technique that helps examine the environment, anticipate changes, and guide decision-making ([Zettelkasten Evolution - A Wardley Map : r/wardleymaps - Reddit](#)). By applying this framework, we chart how an EU-centric cloud can evolve from existing technologies (commodity components) toward a robust, independent ecosystem, and identify where open-source efforts can best accelerate progress.

## Introduction

**Context – Digital Sovereignty Challenges:** There are growing concerns in Europe over digital sovereignty and the heavy reliance on U.S.-based cloud providers (Amazon Web Services, Microsoft Azure, Google Cloud). European governments and industries worry that dependence on foreign cloud infrastructure could expose them to data privacy risks and geopolitical vulnerabilities. For example, the U.S. CLOUD Act allows American authorities to access data held by U.S. tech companies even if stored in EU data centers ([US Cloud Act: Threat for European Data Protection](#)) ([US Cloud Act: Threat for European Data Protection](#)), directly clashing with EU privacy regulations. This has accelerated calls within the EU for home-grown cloud solutions that keep data under European jurisdiction and control ([Cloud Clash: Europe Divides Over Data Digital Sovereignty - CEPA](#)).

**Challenges – Toward a Sovereign Cloud Model:** The key challenge is building a cost-effective, interoperable, and decentralized cloud model that doesn't force a complete overhaul of existing IT infrastructure. European businesses have invested heavily in the major public clouds, and moving to a new platform is often *fraught with difficulty* – current “sovereign cloud” options usually require **significant investments in new resources for integration, testing and interoperability** ([EU Sovereign Cloud Initiative Drives Single-Source Solutions | Business Wire](#)). Thus, any EU sovereign cloud must offer a *low-friction migration path* for users. Ideally, organizations should be able to port their applications with minimal changes, preserving their existing workflows and tools. Another challenge is ensuring decentralization: instead of one monolithic provider, the model should allow deployments at EU-wide, national, or even municipal levels to improve resilience and local control. Finally, cost-effectiveness is crucial – a solution built on open-source and commodity hardware can avoid the expense of proprietary systems, making sovereignty economically feasible.

**Language as an AI Imperative:** An often overlooked aspect of technological sovereignty is **language sovereignty**. English dominates AI research and tools, which means that users and developers who are non-native English speakers face additional barriers ([Towards Truly Multilingual AI: Breaking English Dominance](#)). In the context of large language models (LLMs) and AI assistants, this dominance creates a risk that European languages (especially those less represented online) could be left behind. Portuguese is a prime example: it's one of the world's most spoken languages,

spanning Europe, South America, Africa, and beyond ([World Portuguese Language Day | UNESCO](#)), yet AI systems trained primarily on English may not serve Portuguese speakers equally well. The EU's commitment to cultural and linguistic diversity implies that its AI infrastructure should natively support interaction in languages like Portuguese, French, German, etc. This is not just a cultural issue but a technical one — AI models perform better in languages with abundant training data. Without deliberate effort, the gap between English and other languages in AI capabilities will widen, leading to a **linguistic divide** where non-English users get inferior outcomes ([ChatGPT Is Cutting Non-English Languages Out of the AI Revolution | WIRED](#)). Therefore, integrating European languages into AI development (data, models, interfaces) from the ground up is essential for an inclusive digital future.

This white paper addresses these twin priorities: building a **sovereign EU cloud infrastructure** that is open and compatible, and ensuring **AI language inclusivity** so that technologies serve all European communities equitably. We outline a technical and strategic approach for the cloud stack, and discuss why and how Portuguese (as a representative case) and other European languages should be embedded in the AI ecosystem. We also provide policy and industry recommendations to drive these initiatives forward.

## Technical and Strategic Approach to EU Cloud Infrastructure

### Open-Source Cloud Stack

A cornerstone of EU's sovereign cloud should be a fully **open-source cloud stack**. Embracing open-source software ensures transparency (the code can be inspected for security backdoors or compliance), and it prevents lock-in to any single vendor. In fact, **only open source guarantees digital sovereignty by interoperability, transparency and independence from unauthorized interference** ([Sovereign Cloud Stack](#)). By using open frameworks, the EU can avoid reliance on proprietary technologies controlled by foreign entities.

Fortunately, we do not need to start from scratch. The strategy is to leverage proven open-source components from the cloud-native ecosystem and integrate them into a cohesive stack. For example, the Sovereign Cloud Stack (SCS) initiative is already **using and standardizing existing open source components such as Kubernetes**, extending them where required ([Sovereign Cloud Stack](#)). Kubernetes provides a foundation for container orchestration, OpenStack or CloudStack can handle infrastructure-as-a-service, and other projects can cover storage, networking, and identity management — all under open licenses. The result is a modular cloud platform that any provider or government can deploy. By sharing best practices

for operating these open tools, even smaller players (SMEs or municipalities) can offer high-quality cloud services without proprietary dependencies ([Sovereign Cloud Stack](#)).

A practical example of the power of open-source cloud emulation is **LocalStack**. LocalStack is a cloud service emulator that runs locally, providing a sandbox that mimics the AWS cloud's functionality ([GitHub - localstack/localstack: A fully functional local AWS cloud stack. Develop and test your cloud & Serverless apps offline](#)). Developers use it as a “drop-in replacement” for AWS in development and testing, running AWS APIs on their own machines. This concept can be extended to the sovereign cloud: if a local laptop can emulate AWS services, a network of servers certainly can. In essence, technologies like LocalStack demonstrate that a fully transparent re-implementation of cloud APIs is feasible. (Notably, LocalStack already **emulates over 90 AWS services** in functionality ([Accelerating software delivery using LocalStack Cloud Emulator from AWS Marketplace | AWS Marketplace](#)).) By basing the EU cloud on open-source emulations or reimplementations of cloud APIs, we ensure that the platform's behavior is predictable and verifiable. Moreover, an open-source cloud stack invites contributions from the global community – bugs can be fixed and features added by a wide range of contributors, accelerating innovation. Open-source development is also aligned with European values of collaboration and openness, and it can harness talent from universities, startups, and IT companies across EU member states.

## Seamless API Compatibility

To achieve broad adoption, the EU sovereign cloud must offer **seamless API compatibility** with existing popular cloud platforms. This means an application built to use Amazon S3 for storage or Google's Pub/Sub for messaging could be pointed to the EU cloud equivalent and run with little to no modification. High fidelity in API compatibility is crucial because it **dramatically lowers the barrier to migration**. One of the reasons enterprises hesitate to leave the big cloud providers is the potential cost and complexity of re-engineering their applications for a new environment ([EU Sovereign Cloud Initiative Drives Single-Source Solutions | Business Wire](#)). If the sovereign cloud mirrors the APIs of AWS/Azure/GCP, companies can migrate by simply redirecting their cloud endpoints, avoiding massive rewrites or retraining of staff.

Implementing this compatibility is ambitious but aided by the open-source community. There have been past efforts like Eucalyptus (which mimicked AWS EC2/S3 APIs) and current projects like OpenStack that offer an alternative API. The sovereign cloud can incorporate **API translation layers or compatibility modules** that accept calls in the AWS or Azure format and execute them with the underlying open-source services. In effect, it would “speak” the same language as the major clouds. We already see hints of this approach: for instance, **LibreGraph provides a sovereign API fully compatible with the Microsoft Graph API** to replace Microsoft's cloud services in a privacy-respecting manner ([LibreGraph: An open cloud API for](#)

[privacy and sovereignty](#)). Similarly, LocalStack's success in reproducing AWS behaviors locally (covering the majority of AWS services ([Accelerating software delivery using LocalStack Cloud Emulator from AWS Marketplace | AWS Marketplace](#))) indicates that near-total API compatibility is achievable with sufficient effort.

This approach would allow European organizations to **migrate to a sovereign cloud with minimal overhead** – a critical requirement given that compatibility issues have been a major pain point in early sovereignty efforts ([EU Sovereign Cloud Initiative Drives Single-Source Solutions | Business Wire](#)). A developer should be able to take an existing cloud automation script (Terraform templates, Kubernetes deployments, etc.), switch out the endpoints to the EU cloud, and deploy successfully. By maintaining **parity with existing cloud APIs**, the EU cloud also benefits from existing knowledge and tooling. DevOps teams can apply their current skills without steep learning curves, and popular open-source tools (CI/CD pipelines, infrastructure-as-code frameworks, monitoring dashboards) will work out-of-the-box. This greatly enhances the usability of the sovereign stack and encourages adoption.

## Decentralized and Multi-Level Deployment

Central to European cloud sovereignty is **decentralization**. Rather than a single, monolithic cloud provider, the vision is a *federation* of cloud providers and deployments at multiple levels (EU-wide, national, regional, municipal). The cloud stack should be designed such that it can be deployed in a distributed fashion, with interoperability baked in. In practice, this means a company or government agency could choose to run an instance of the cloud on-premises or in a local data center, and still seamlessly connect with other instances across Europe when needed.

Sovereign Cloud Stack's model provides a blueprint: it enables **federated cloud services** where users can leverage distributed cloud resources from several SCS operators ([Sovereign Cloud Stack](#)). European companies thus have options ranging from running the entire cloud stack themselves, to using services from a mix of local providers – achieving *100% freedom of choice* and **avoiding lock-in or dependency on providers from foreign jurisdictions** ([Sovereign Cloud Stack](#)). This federated approach means that, for example, a municipality could host citizen data and applications on its own servers for full control, but still tap into extra capacity or specialized services from a national cloud service when necessary. All deployments, being based on the same open standards, would interoperate smoothly.

Multi-level deployment enhances **resilience and control**. Data can be kept close to its source (satisfying data locality laws and reducing latency for users), and if one node in the federation faces an outage or security threat, others can continue operating independently. It also fosters competition



and innovation: numerous providers (including startups or public institutions) can offer sovereign cloud services, preventing the emergence of a single monopoly. Crucially, decentralization aligns with Europe's political structure – respecting the sovereignty of member states over their data and infrastructure, while enabling collaboration. Even at the sectoral level, we might see specialized clouds (for healthcare, education, etc.) that adhere to the sovereign framework but are tailored for specific regulatory needs.

Technically, achieving a federated cloud involves implementing robust **interoperability and governance mechanisms**. Open standards for identity and access management, data exchange, and service discovery will allow different cloud instances to trust and communicate with each other. The EU could define certification levels (as SCS is doing with *SCS-compatible*, *SCS-compliant*, etc.) to ensure any cloud instance meets baseline requirements for security and compatibility. The result would be a network of sovereign clouds forming a cohesive “cloud of clouds” for Europe. This decentralization ensures that the EU cloud is truly **sovereign at every level** – from local communities up to the pan-European scale.

## Wardley Mapping Analysis

To guide the strategic development of the EU sovereign cloud, this paper employs **Wardley Mapping** principles. Wardley Maps are a strategy tool that visualize the components of a system (in this case, cloud infrastructure and related services) along two dimensions: the value chain and the stage of evolution of each component ([Wardley Maps: Business Analysis, Experimentation & Evolution – Software Requirements Management](#)). Components evolve from genesis (novel, custom-built) to commodity (standardized, utility-like) as technology matures and markets develop. Using this framework helps identify which parts of the cloud stack need innovative development and which parts should be treated as commodities (and thus possibly outsourced or provided via utility).

Applying Wardley Mapping to cloud infrastructure reveals that many basic cloud services (compute, storage, network) are already **commoditized** – they are provided at massive scale by hyperscalers as utility services, and users expect them to just work cheaply and reliably. In Wardley Map terms, cloud computing has largely moved to the rightmost side (commodity/utility) of the evolution axis ([\[PDF\] Wardley Maps - coach-agile.com](#)). This means the EU sovereign cloud effort should not focus on reinventing the wheel for these components; instead, it should *leverage existing mature solutions*. For example, commodity components like servers and hypervisors can be standard open-source implementations (Linux, KVM, etc.), and even higher-level services like object storage or databases can build on open-source projects that are already stable and widely used. The value-add of the EU cloud will not be in creating a new form of VM or database, but in integrating and delivering these in a sovereign way.



On the other hand, Wardley Mapping may highlight areas where innovation (*genesis or custom-built components*) is needed. For instance, a truly seamless API compatibility layer for all major cloud services might be something new that must be developed (since it's not fully provided by any single open project yet). Or innovative governance tools for the federated model might need to be created. These would be on the left side of the map (unknowable or early stage), requiring experimentation and bespoke development.

Crucially, Wardley Maps help in sequencing our strategy. Given that open-source **is an accelerator of evolution** – it helps activities mature toward industrialized forms faster ([Bits or pieces?: Some basics of operation.](#)) – we see that by choosing open-source implementations for as many components as possible, we hasten the commoditization and reliability of the whole stack. Open-source also encourages a competitive market of providers around those commoditized components, which aligns with the EU's goal of avoiding dependency.

In summary, the Wardley Mapping analysis underpins this approach by confirming that **the core cloud services should be treated as utilities** and provided via open-source, commodity solutions (ensuring stability and cost-efficiency), while the differentiation of the EU sovereign cloud will come from how these services are deployed (federation, governance) and integrated (compatibility layers, multilingual support, etc.). This strategic mapping ensures the EU invests effort in the right places: innovating where it must (to achieve sovereignty or inclusivity goals) and not wasting resources duplicating what the market already offers as a commodity. It provides a clear evolutionary roadmap for the sovereign cloud initiative and helps communicate the plan to stakeholders by illustrating where value is created and where reliance on existing components is acceptable.

## The Role of Portuguese and Other European Languages in AI

### Portuguese as a Programming Language in the AI Era

Traditionally, interacting with computers — especially programming — has required using the English language or syntax heavily influenced by English. From programming keywords (if, else, for) to libraries and documentation, English has been the default. In the AI era, however, we have an opportunity to change that paradigm. With advanced AI and natural language interfaces, **natural language itself becomes the programming interface**. This means that one could potentially "write code" by instructing an AI in plain Portuguese, and have it execute tasks or generate software. In other words, Portuguese can effectively become a *programming language* when an AI system understands and acts on it accurately.

The benefit of this shift is enormous for non-English speaking populations. As noted, non-English speakers often face the dual hurdle of learning complex technical skills *and* a foreign language (English) to use them ([Towards Truly Multilingual AI: Breaking English Dominance](#)). By enabling AI systems to be used in Portuguese, we lower the barrier to entry for millions of new programmers, domain experts, and everyday users. For example, a Brazilian data analyst could ask a database question in Portuguese and get results, or a Portuguese doctor could interact with a medical expert system in her native tongue. This goes beyond simple translation; the AI needs to truly understand Portuguese context, idioms, and technical expressions to function reliably.

Implementing Portuguese as a "programming language" in AI involves training models on Portuguese instructions and integrating Portuguese-language processing at every level of the software stack. Recent large language models show promise here – if given sufficient Portuguese training data, models like GPT-4 can interpret and respond to complex prompts in Portuguese. We should push this capability further, developing AI coding assistants that can generate code from Portuguese descriptions of a problem, or configure systems based on Portuguese directives. Imagine describing an app you want to build in Portuguese, and the AI produces a prototype. That kind of accessibility democratizes technology creation.

This effort also has symbolic importance: it affirms that **English need not be the sole language of technology**. Just as the EU has 24 official languages and conducts business in all of them, our AI systems should recognize and operate in those languages. Portuguese, given its widespread use globally, is an excellent candidate to lead this change. By expanding the domains in which Portuguese is used (from daily conversation into cutting-edge AI interactions), we ensure the language stays vibrant and relevant in the digital age. This is a case of technology adapting to people, rather than forcing people to adapt to technology.

## AI Bias and the Linguistic Divide

One of the risks of the current AI landscape is the emergence of a **linguistic divide** – a scenario where AI works excellently for English (and perhaps a handful of other languages), but performs poorly for others. Unfortunately, evidence of this divide is already visible. AI chatbots and language models are **significantly less capable in languages other than English** ([ChatGPT Is Cutting Non-English Languages Out of the AI Revolution | WIRED](#)). This stems from the fact that the largest AI models are trained on predominantly English datasets, or otherwise give priority to English because that's where the data is most abundant. The dominance of English (and to a lesser extent, languages like Chinese) in AI training data **compounds the bias in AI systems** ([AI Language Revolution: addressing bias, multilingual challenges, and workforce impact](#)) – they inherently become better at those languages, creating a self-reinforcing cycle of AI serving the already digitally-dominant languages.

For speakers of Portuguese and many other languages, this can mean AI applications that misunderstand queries, produce flawed results, or simply aren't available in their language. If an AI assistant is far less fluent in Portuguese than in English, Portuguese speakers are effectively second-class citizens in the AI revolution. This is not just inconvenient; it can be economically and culturally damaging. Think of global commerce or innovation – if AI tools give English-speaking businesses a much better service (say, more accurate customer analytics or smarter chatbots) than those available to Portuguese-speaking businesses, the latter are at a competitive disadvantage ([ChatGPT Is Cutting Non-English Languages Out of the AI Revolution | WIRED](#)). Moreover, important cultural and linguistic nuances might be lost or misinterpreted by AI that wasn't trained to understand them, leading to outputs that are irrelevant or even offensive in a local context.

There is also the issue of **cultural bias**. AI models reflect the data they are trained on. An English-centric AI may inadvertently carry Anglo-American cultural assumptions that do not hold in Europe or elsewhere. Pascal Fung, an AI researcher, pointed out that if we don't address these biases, AI will *reinforce the primacy of English* and the perspectives tied to it, rather than challenging the status quo ([ChatGPT Is Cutting Non-English Languages Out of the AI Revolution | WIRED](#)) ([Towards Truly Multilingual AI: Breaking English Dominance](#)). This could marginalize knowledge and content in other languages. For example, valuable research written in Portuguese might be ignored by AI because it wasn't in the training set, or a conversational AI might fail to use polite forms of address that are important in European languages.

To bridge this linguistic divide, we need **localized AI development**. This means curating large Portuguese (and other language) datasets, training models in those languages, and rigorously evaluating AI performance across languages. It also means building AI products with multilingual support from the outset, not as an afterthought. As a policy, any critical AI system deployed in Europe should be tested for how well it works in all official languages. If ChatGPT-like models or other AI services are less fluent in Portuguese, that gap must be seen as a bias to be fixed, not an acceptable norm. By shining a light on these issues now (the early days of the AI revolution), the EU can take steps to ensure **non-English speakers are not left behind** ([ChatGPT bias: 3 ways non-English speakers are being left behind](#)). This is analogous to web accessibility – just as we believe websites should be accessible to people with disabilities, AI should be accessible to people in *their native languages*.

In short, acknowledging and addressing AI's linguistic bias is critical to prevent new inequalities. The goal must be AI that understands and serves all Europeans equally well, whether they speak Portuguese, Polish, or Finnish. The next section discusses how we can achieve parity for these languages in AI systems.

## Developing AI Language Parity

How can we ensure that AI models treat Portuguese (and other European languages) as first-class citizens? Achieving **language parity** in AI requires action on multiple fronts: data, technology, and policy.

From a **data perspective**, we need rich, high-quality datasets in Portuguese and other underrepresented languages. This includes not only general web data but specialized data (medical text, legal documents, literature, conversational data, etc.) to cover the full breadth of language use. Initiatives should be undertaken to create and open-source such datasets. For instance, collecting a large Portuguese text corpus for training language models, or a speech dataset for speech recognition in Portuguese. The European Commission and national governments can fund projects to compile these resources. There has been progress: the European Language Equality (ELE) project has developed a strategic agenda aiming for **full digital language equality in Europe by 2030** ([Strategic Agenda - European Language Equality](#)), highlighting the importance of comprehensive language resources and tools.

On the **technology side**, we must train and fine-tune AI models specifically for these languages. A promising development is the **OpenEuroLLM project**, which is creating the first family of open-source large language models covering all official EU languages ([A pioneering AI project awarded for opening Large Language Models to European languages | Shaping Europe's digital future](#)). With EU funding, OpenEuroLLM brings together startups and research labs to ensure that languages like Portuguese, Czech, and Greek have dedicated AI models, not just translated versions of English models. By training on European multilingual data and using European supercomputers, they aim to produce AI that inherently understands the linguistic nuances of each language. Such models could be used as foundational building blocks for various applications – from chatbots that converse in local languages to machine translation systems between European languages that are more accurate than generic ones.

Beyond large language models, parity means focusing on **evaluation and optimization**. AI models should be benchmarked in all target languages. If a model answers questions correctly 90% of the time in English but only 70% in Portuguese, developers need to identify why and address it (perhaps by training on more Portuguese data or adjusting the model architecture). Likewise, voice assistants should recognize accented Portuguese from Portugal and Brazil with equal proficiency as they do English from America. This requires iterative testing and improvement.

Another strategy is encouraging the development of **community-driven language models**. Just as open-source is key for cloud sovereignty, open collaborative development is key for language tech. Portuguese-speaking AI researchers and enthusiasts could be supported to build language models (some communities have done this, for example, there are open-source Portuguese NLP libraries and models trained by academia). These community models can be integrated or distilled into larger systems, ensuring that progress in Portuguese NLP keeps pace with English.

From a **policy and industry standpoint**, setting requirements for multilingual support in AI products can drive change. For example, public sector tenders for AI solutions in the EU could require bidders to support at least a certain level of functionality in all EU official languages. This creates a business incentive for companies to invest in that support. The private sector in Portuguese-speaking countries (like Portugal and Brazil) can also collaborate – sharing data and techniques to improve Portuguese AI. There is a natural synergy here: Brazil's large tech community and Portugal's EU context together can strengthen Portuguese-language AI tools that benefit all Lusophone communities.

Finally, we should mention **education and awareness**: developers and researchers need to be aware of language bias issues. Including courses on multilingual AI in university curricula, and encouraging research publications in languages other than English (with translation for broader consumption), can help normalize a multilingual approach in AI development. If the upcoming generation of AI practitioners considers it normal to work in multiple languages, parity will be much closer to reality.

In conclusion, achieving AI language parity is a multi-faceted endeavor. By investing in data, fostering open-source multilingual models (like OpenEuroLLM), and aligning industry incentives with multilingual performance, we can ensure that AI systems serve Portuguese speakers just as well as they serve English speakers. This is crucial for fairness, cultural preservation, and tapping into the full talent pool of AI users and developers across Europe.

## Policy and Industry Recommendations

### Government & EU Initiatives

European governments and the EU institutions should take an active role in realizing cloud sovereignty and AI language inclusivity. **Policy support and funding** are key levers. A prime example is Germany's funding of the Sovereign Cloud Stack project – the German Federal Ministry of Economic Affairs and Climate Action explicitly backed SCS to bring Gaia-X's vision to life by developing an open-source cloud foundation ([Sovereign Cloud Stack](#)). Similarly, the EU can use programs like Digital Europe, Horizon Europe, or the new Strategic Technologies for Europe Platform (STEP) to fund the development of the sovereign cloud stack and associated tools. Grant funding can accelerate the creation of API-compatible services or the federated interoperability mechanisms discussed earlier.

In addition to funding R&D, **government procurement** can drive adoption. If EU agencies and national governments start preferring the open sovereign cloud for their own needs, it provides an early market and signals confidence. Public sector demand can also encourage local IT providers to offer sovereign cloud services, knowing there's a client base. On the regulatory side, the EU could establish a **cloud certification (or compliance) framework** defining what constitutes a “sovereign cloud” (as hinted by discussions around the EU Cloud Rulebook or similar initiatives). This framework can set requirements for data residency, open-source usage, interoperability, and security standards. Having clear definitions will guide industry efforts and reassure users of what they're getting ([Key Insights from the EU Sovereign Cloud Day in Brussels, 2024](#)).

For AI language inclusivity, governments should embed multilingual requirements into AI strategies. The EU's AI Act, for instance, could include provisions about transparency and performance across languages for certain AI systems deployed in public contexts. The EU has already demonstrated foresight by funding projects like **OpenEuroLLM** to ensure European languages are covered by cutting-edge models ([A pioneering AI project awarded for opening Large Language Models to European languages | Shaping Europe's digital future](#)). Continuing and expanding such programs is recommended – perhaps a dedicated **Multilingual AI Fund** that supports resources and startups focused on language technology for Europe. Also, pan-European collaborations (involving institutions like the European Language Grid or CLARIN) can share language resources and avoid duplication of efforts across countries.

Another important aspect is international cooperation within the Portuguese-speaking world. Portugal can collaborate with Brazil, Angola, Mozambique, and other Lusophone countries on AI language data and research. An EU-backed initiative to strengthen AI in Portuguese would not only benefit EU citizens but also enhance ties with those nations (an example of tech diplomacy). This could take the form of joint research centers or student exchange programs focusing on Portuguese NLP and speech technology.

In summary, **public sector leadership** is essential. By providing funding, setting standards, and acting as early adopters, European governments and the EU can jump-start the sovereign cloud and ensure AI language inclusivity. These actions de-risk the projects for private players and align the ecosystem toward common goals.

## Private Sector Engagement

The vision outlined in this paper cannot be achieved by the public sector alone; **industry involvement is crucial**. European cloud providers, software companies, startups, and even enterprises as end-users should all be mobilized to contribute to sovereign cloud and multilingual AI initiatives.



For the cloud infrastructure, existing European cloud companies like OVHcloud, Deutsche Telekom (T-Systems), Orange Business Services, and others should see the open-source sovereign stack not as a threat, but as an opportunity. By pooling resources to build a common infrastructure (on which they can differentiate with value-added services or better customer support), they can collectively compete with the scale of U.S. hyperscalers. In fact, many companies have already come together under coalitions such as the **Sovereign Cloud Stack community and the OSB Alliance**, where **numerous companies joined forces to standardize and certify an open cloud stack**. This kind of cooperation should be expanded and sustained. Industry players can contribute engineering effort to the open-source projects, ensuring the features they need are implemented. They can also form support ecosystems around the sovereign cloud (similar to how companies offer support/services for pure open-source projects like Linux). The Gaia-X association, with its many member companies, is another venue where the private sector can coordinate requirements and ensure interoperability around open, sovereign solutions ([EU Sovereign Cloud Initiative Drives Single-Source Solutions | Business Wire](#)).

Startups and innovators should be encouraged to build on top of the sovereign cloud platform. If the API-compatibility is strong, any tool made for AWS could potentially be adapted to run on the EU cloud – this is enticing for European startups who currently might depend on AWS/Azure and worry about costs or data location. The EU could consider incubator programs or challenges for startups to create services (AI platforms, IoT backends, etc.) specifically optimized for the sovereign cloud. This creates a virtuous cycle: the more services and tools available on the platform, the more attractive it becomes to other users, and so on.

Regarding AI and language, private sector players in Europe's AI industry need to prioritize multilingual support as a market differentiator. We already see language service providers and AI firms acknowledging that non-English AI is a huge untapped market ([AI Language Revolution: addressing bias, multilingual challenges, and workforce impact](#)). Companies like voice assistant manufacturers, chatbot developers, and software vendors should invest in Portuguese-language support knowing there is a large user base that prefers Portuguese. For example, a customer service AI startup could ensure its product can handle Portuguese out-of-the-box, giving it an edge in Portugal, Brazil, and parts of Africa. Tech giants and AI labs should open-source or share more of their multilingual models to spur adoption (the way Meta open-sourced their “No Language Left Behind” translation model is a positive example).

Moreover, industry can help with **data sharing partnerships**. Telecom companies or media companies in Portuguese-speaking countries might have datasets (like voice recordings or text transcripts) that can aid AI model training. With proper privacy safeguards, these could be made available to researchers or included in the open datasets mentioned earlier.



Finally, it's worth noting that supporting these initiatives can be good PR and align with corporate social responsibility goals. A company that actively supports Europe's digital sovereignty and linguistic diversity can brand itself as a champion of user rights and cultural preservation. This could resonate well with European customers and governments.

In conclusion, the private sector should embrace the sovereign cloud and multilingual AI not just as compliance exercises, but as strategic opportunities. By collaborating on foundational infrastructure and then competing on services, they can create a rich cloud ecosystem independent of Big Tech dominance. And by baking multilingual capabilities into products, they can reach underserved markets and improve user satisfaction. The EU can facilitate this through public-private partnerships, but ultimately the creativity and efficiency of the private sector will drive these solutions to maturity.

## Education and Workforce Development

No transformation is complete without people. To support a sovereign cloud and multilingual AI, Europe needs to cultivate a skilled **workforce** that can build, maintain, and innovate these systems. This calls for updates in education and professional training.

**Curriculum for Cloud Sovereignty:** Universities and technical institutes across Europe should incorporate cloud computing courses that focus on open-source technologies and interoperability. Courses on distributed systems can include modules on OpenStack, Kubernetes, and tools like Terraform – giving students hands-on experience with the very components that make up the sovereign cloud stack. Additionally, case studies on digital sovereignty (covering Gaia-X, SCS, etc.) can be introduced in IT policy or business courses, so the next generation of decision-makers understands the importance of these issues. There may also be value in specialized programs or certifications (potentially an EU-backed certification) for “Sovereign Cloud Engineer” or “Cloud Interoperability Specialist,” which signal expertise in deploying and managing clouds in a vendor-agnostic way.

**Multilingual AI in Education:** For AI and data science programs, multilingualism should be part of the training. This could involve teaching NLP (Natural Language Processing) with examples in multiple languages, not just English. Students should learn about language models for French, Portuguese, German, etc., and the challenges of low-resource languages. Competitions or hackathons can be organized where students develop an AI solution that works in several European languages. Such exercises would normalize the idea that an AI developer in Europe is expected to think

beyond English. Supporting educational content in local languages (like textbooks or online courses for AI in Portuguese) will also help—learning complex subjects in one’s native language can be more effective and will attract more talent into the field.

**Reskilling and Upskilling:** For the current workforce, especially IT professionals, there should be opportunities to gain skills in these new focus areas. Cloud administrators proficient in AWS/Azure could be offered crash courses or workshops on the open-source equivalents. Governments and industry can partner on **training programs** to help IT staff get familiar with the sovereign cloud stack deployment and troubleshooting. This is similar to how, in the past, training programs helped people transition from proprietary software to open-source alternatives in government. On the AI front, existing software developers might need training on incorporating language support in their apps, or how to fine-tune an open-source language model for their domain.

**Community and Knowledge Sharing:** Europe could foster communities of practice around these topics. For example, an online forum or annual conference for “EU Sovereign Cloud & AI” where practitioners share experiences, tools, and lessons learned. This would help diffuse knowledge quickly. Perhaps the EU or academic institutions can host sandboxes or labs where people can play with a mini-deployment of the sovereign cloud, or experiment with training models in various languages, without needing their own heavy infrastructure (leveraging European supercomputers for public research access, as is being done for some AI projects).

**Inclusion of Linguists and Social Scientists:** It’s also worth noting that developing multilingual AI isn’t just a technical challenge; it benefits from linguistic and cultural expertise. So, education efforts should be interdisciplinary. Programs that combine computer science with linguistics (computational linguistics programs) should be supported and expanded. Likewise, policy programs could combine technology with language policy education, to create experts who can navigate the intersection of AI and multilingual governance.

By investing in people, Europe ensures that it has the autonomy not just in technology, but in the know-how to evolve and sustain that technology. A well-trained workforce can adapt open-source tools to local needs, ensure security, and innovate new features. And when it comes to AI, developers who are themselves diverse (speaking multiple languages, coming from different cultures) will build more inclusive systems. This human capital aspect is the foundation upon which the lofty goals of sovereign infrastructure and inclusive AI will be built.

## Conclusion

The trajectory for Europe's digital future can be one of **self-determination and inclusivity**. This white paper has argued that the EU's sovereign cloud must be **open, compatible, and decentralized** at its core. Openness provides transparency and freedom from undue influence, compatibility ensures that Europeans can transition to sovereign tech without prohibitive costs, and decentralization guarantees resilience and alignment with Europe's federated structure. In practical terms, this means leveraging open-source for everything – from infrastructure to APIs – and designing the system such that any European entity can adopt it and interoperate, thereby achieving *100% freedom of choice and no lock-in to foreign providers* ([Sovereign Cloud Stack](#)).

Simultaneously, the evolution of AI in Europe must prioritize **multilingual, inclusive, and equitable** development. It is not enough for AI to be cutting-edge; it must also be culturally and linguistically aware. European languages like Portuguese should remain integral to our technological progress, not sidelined by an English-first approach. The vision is for a future where a Portuguese researcher, a Spanish doctor, or a French engineer can use AI tools in their native language with the same efficacy as an English speaker. Achieving this will ensure that AI serves to **unite and empower Europe's diverse populations, rather than create new divides** ([Towards Truly Multilingual AI: Breaking English Dominance](#)).

There is a strategic and economic opportunity in aligning these two endeavors – cloud sovereignty and AI language diversity. By investing in both, the EU and its partners (including Portuguese-speaking nations globally) can define a unique value proposition on the world stage. They can offer a model of technology that respects local autonomy and cultural identity, something increasingly attractive in a world wary of one-size-fits-all solutions from Big Tech. For instance, a sovereign EU cloud that natively supports Portuguese could be an appealing solution not only in Europe but also for Brazil's public sector or Angola's businesses, strengthening international ties through technology. In effect, Europe can export a vision of **tech sovereignty + multilingual AI** as an alternative path of innovation – one that prioritizes ethical standards, user rights, and global diversity.

In closing, the path forward requires continued **commitment and collaboration**. European policymakers must set the direction and create enabling conditions. Technologists and open-source communities must build and iterate on the cloud and AI tools. Industry must drive adoption and contribute expertise. And educators must prepare the next generation to carry this mantle. If all stakeholders move in concert, the EU can achieve a cloud infrastructure that is truly its own and an AI ecosystem that speaks *all* the languages of its people. This will safeguard Europe's digital sovereignty and ensure that in the AI era, no language or community gets left behind – a future where technology and linguistic heritage grow hand in hand.