

Sovereign AI Index: Methodology & Inclusion Criteria

Center for a New American Security

April 20, 2026

Executive Summary

The Sovereign AI Index, developed by the Technology and National Security Program at the Center for a New American Security (CNAS), tracks national projects to build sovereign AI capabilities. The index focuses on government-backed projects to proactively invest in domestic AI infrastructure, models, and data ecosystems to reduce foreign dependency and advance sovereign interests.

1) Purpose and Scope

What this index tracks: This database catalogs substantive sovereign AI projects worldwide, excluding those from the United States and China. The United States and China may be added in future versions, but the primary goal of the index is to track sovereign AI deployments beyond the two AI leaders. The index includes the following projects when there is a government role, as further defined in this document:

- 1. Infrastructure:** Compute infrastructure programs, such as graphics processing unit (GPU) clusters and AI data centers
- 2. Models:** AI models, including large language models (LLMs)
- 3. Data:** Initiatives to organize and harness national data to build AI capability

Where applicable, the index also includes government programs that fund or provide access to the above, as well as projects that combine elements across categories.

2) Inclusion Criteria

Core requirements: For inclusion in the index, projects must meet all three of the following criteria:

- 1) Material government support:** Projects must show both a documented government instrument and a form of material contribution.
 - a) Instruments:** legislation or regulation; official strategy or announcement naming the project; memorandum of understanding (MOU) or intergovernmental/public-private agreement; procurement contract, tender award, or grant agreement
 - b) Contributions:** direct public funding (e.g., appropriations, grants, cost sharing); sovereign wealth fund or state-owned enterprise investment; project-linked incentives (land grants, concessions, tax credits); nonmonetary government resourcing (dedicated compute allocations, institutional infrastructure, operation by a government-funded institution)

Note: The index does not include verbal endorsements or “strategic partnership” announcements in press releases. Without both a documented instrument and contribution, such statements reflect aspiration rather than meaningful government involvement.

- 2) Sovereign intent:** Projects must have an explicit objective to treat domestic AI capacity as a strategic asset to protect or advance at least one of the following national interests:
 - a) National security
 - b) Economic opportunity, competitiveness, and resilience
 - c) Cultural and linguistic preservation and fidelity
- 3) Substantive AI deployment:** Projects must involve a nontrivial public investment or in-kind contribution (see “Investment thresholds” below) in at least one of the following:
 - a) Infrastructure: AI data centers, GPU- or AI-optimized supercomputers, or AI compute clusters, including compute access programs that allocate or

subsidize AI compute for domestic users (even if the underlying hardware is vendor operated). These categories may overlap.

b) Model: Foundation model training or adaptation of pretrained models (fine-tuning, continued pretraining, instruction tuning, or comparable methods)

c) Data: Curated datasets or language corpora intended for AI development, training, evaluation, or deployment

Investment thresholds: The index evaluates whether a project represents a substantive public investment using one of two approaches, depending on whether the project's budget has been publicly disclosed.

- Projects with disclosed budgets. Government or intergovernmental contributions must meet or exceed \$2 million for infrastructure projects or \$250,000 for model and data projects. The lower floor for model and data projects reflects the lower cost of model adaptation and dataset curation relative to physical infrastructure. Both floors are intentionally low: Sovereign AI deployments occur at a modest scale in countries with limited public investment capacity, and excluding such projects would conflate a nation's financial capacity with its sovereign ambition.
- Projects without disclosed budgets. Government disclosure practices vary substantially across jurisdictions, and a significant share of sovereign AI activity occurs in contexts where project-level budgets are not publicly itemized. For these projects, the research team assesses inclusion based on two factors: (1) a documented mechanism by which government capital reaches the project, such as a named compute allocation, operation by a government-funded institution, or inclusion in a disclosed program budget; and (2) scale indicators—hardware, model size, dataset scope, or operational reach—consistent with

substantive sovereign AI activity.¹

Hosting requirement: For infrastructure and compute access projects, AI compute must be physically hosted in-country or in-region (for regional programs such as the European High Performance Computing [EuroHPC] AI Factory Antenna Program). For model and data projects, in-country hosting is not required, provided public funding is directly tied to the model or data effort and sovereign intent is explicit.

Time frame: Projects must have been announced or become active on or after December 1, 2022, a cutoff date chosen to capture the wave of government AI investment catalyzed by the release of ChatGPT the prior month. Preexisting facilities are included only if they received AI-specific upgrades, repurposing, or new investment after December 1, 2022.

3) Projects That Do Not Meet Inclusion Criteria

To maintain focus on genuine sovereign AI capability-building projects, the index excludes the following:

- a) Pure regulatory or governance frameworks without technology investment
- b) Enterprise AI initiatives advertised as promoting sovereignty without government backing
- c) Generic digital transformation strategies without specific AI sovereignty components
- d) Research-specific funding instruments (including grants), unless they directly fund a national capability-building project (e.g., dedicated compute, a national dataset/data platform, or a government-backed model program with public access or service obligations)

¹ This approach applies different evidence standards to disclosed-budget and opaque-budget projects. The index takes this tradeoff deliberately: applying disclosure-contingent thresholds uniformly would systematically underrepresent sovereign AI activity in jurisdictions where budgets are not published. The research team's view is that comprehensiveness with documented judgment serves readers better than uniform thresholds that filter on reporting practices rather than substantive sovereign AI activity. The index will continue to refine its methodology as sovereign AI matures as a policy category, as government disclosure practices evolve, and as the research team accumulates more comparative evidence across jurisdictions.

- e) Cloud or other technology projects without a substantial role for AI, even if they are marketed otherwise
 - f) Government contracts that primarily procure off-the-shelf commercial AI access—such as application programming interfaces (APIs), software-as-a-service offerings, or managed services for routine government use even when fine-tuned for specific use cases—where the procurement does not materially contribute to domestic AI capability.
-

4) Edge Cases

The subsequent rules apply the core criteria to edge cases for which labels or contracting structures make it unclear whether a project is building sovereign capability.

1) Model and Data Programs

- a) Operates or administers government-funded GPU/AI compute for national access
- b) Stewards a government-backed national dataset with explicit sovereignty objectives
- c) Leads or funds the training or fine-tuning of a model explicitly designed to serve national languages, contexts, or public institutions

2. Subnational and Supranational Projects

Sovereign AI capability does not always map neatly onto nation-states. Regional consortia pool resources across borders to achieve scale that individual countries cannot, while subnational governments sometimes act as de facto sovereign AI actors within their own jurisdictions. Excluding these programs entirely would miss a meaningful share of sovereign AI activity. The index accounts for this complexity by distinguishing three types of entries:

- a) Regional umbrella:** Regional programs shared across borders (e.g., OpenEuroLLM) appear as a single entry.

- b) Member-state nodes:** Countries that host a specific facility or contribute distinct national resources to a wider program (e.g., Finland’s LUMI project) receive separate national entries.
- c) Subnational initiatives:** State, provincial, or regional projects that function as independent sovereign AI initiatives for their jurisdiction (e.g., Spain-Catalonia/Aina) also receive separate entries.

3. Foreign Vendor Partnerships

The index includes foreign vendor partnerships only when (a) a national ministry, sovereign wealth fund, public research and development agency, or state-owned enterprise is a contracting party, coinvestor, or direct funder of the AI-specific deployment, and (b) the partnership delivers localized AI-specific infrastructure or capabilities (e.g., GPU clusters, AI supercomputers, national LLM training or fine-tuning, or a government-backed national dataset), as opposed to generic cloud services.

5) Project Status Definitions

Status	Definition	Indicators
Announced	Project has been formally announced	Government announcement backed by documented forms of public commitment via legislation, regulation, MOUs, or other formal intergovernmental or public-private agreements without an active tender

In procurement	Project is in the contracting or acquisition phase	Active tenders or requests for proposals issued, vendor selection in progress, funding allocated
In development	Project is under active construction or development	Contracts signed, construction underway, active development
Operational	Project is live and functioning	Services are at least partially available

6) Data Fields & Definitions

Field	Description	Notes
Country	Name of the project's home country	Lists European Union member states individually; lists supranational projects under the relevant multinational organization's name, such as "European Union" or "United Nations"
Project Name	Project's official name or common designation	Includes acronyms when widely used, but aims to provide the official and complete name
Project Category	Project's primary focus area	Articulates if a project is infrastructure, model, data, or a combination thereof (e.g., model & data). If a project can be subcategorized, the index clarifies this

		distinction in a parenthetical next to the primary focus area. For example: Infrastructure (compute access program)
Purpose	Project’s primary technical function and strategic objective	Focuses on the “what” and “why,” highlighting the specific sovereign value proposition, such as national security, cultural and linguistic preservation and fidelity, and economic competitiveness
Project Milestones	Project’s key dates	Includes the project’s announcement date and status, where determinable (see “Project Status Definitions”)
Location	Project’s physical hosting details: the hosting facility (“Residency”) and the facility’s physical location (“Physical location”)	Identifies a specific hosting facility when one is disclosed; physical location records city, state/province/region, and country. “Not specified” is noted for any element not publicly available
Technical Specifications	Quantitative technical characteristics that enable cross-project comparison of AI capability. Records capacity, architecture, and performance. If undisclosed, states “not specified”	Provides the following specifications by project type: Infrastructure (compute): System type, system architecture, nodes, accelerators, central processing units (CPUs), interconnect, memory, storage, performance, and quantified power/cooling Models: Build type/lineage, parameter, primary languages, training data (tokens + composition), training compute, modalities, and technical

access method²

Data projects: Asset type, scale/volume, modalities, metadata/standards (if stated), update/versioning, and technical access method

Compute-access programs: Allocated resources (e.g., GPU-hours, credits, capacity, model API access), quantitative limits (caps, quotas, duration), and in-scope resources/providers/systems if specified

Access and Licensing

Project access and model or data licensing approach

Specifies the legal release framework for AI models and data (e.g., open source, open weights, proprietary) and the eligibility criteria for infrastructure computing resources (e.g., merit-based access, government-only, startup subsidy)

Budget

Project budget data from official sources or credible media reports

Includes currency as stated in source document. Provides a USD equivalent using the International Monetary Fund monthly exchange rate for the day indicated in the “Budget Date” field, or the first day of the month for which an exchange rate is available when only the month and year are known. States “not specified” if not publicly disclosed

² If a project trains new base weights, it is labeled “Trained from scratch.” If the primary disclosed modification to an existing model is compression through distillation or other methods, it is labeled “Compressed.” Otherwise, it is labeled (1) “Fine-tuned + reinforcement learning” if reinforcement learning is described, (2) “Continued pretraining + fine-tuned” if additional raw-text training is described, (3) “Fine-tuned” if supervised adaptation is described, or (4) “Deployed” if none of these are disclosed. For all model entries not trained from scratch, the base model must be stated in the “Build type/lineage” section. If the base model is not publicly disclosed, it is labeled “Base: Not specified.”

USD Conversion Date	Date of the exchange rate used to convert local currency to USD equivalent	Derives exact dates from official government or corporate sources and approximate dates from credible media reports
Lead Sponsor	Project’s primary government entity or state-affiliated organization providing the mandate or funding	Identifies the government ministry, department, agency, or state-affiliated entity (including sovereign wealth funds and state-owned/affiliated enterprises) responsible for the project’s strategic oversight or material support. For public-private partnerships led by private firms, the public-sector partner is recorded here, and the private firm is listed under “Partners”
Partners	Third-party organization that provides essential inputs required to fulfill the project’s strategic mission that are not possessed or provided directly by the government agency	Lists each partner by name, role (e.g., hardware provider, training, deployment, compute, data), and country of incorporation. Includes the lead sponsor as well Partners include any entity whose technology represents a critical dependency for the project, whether or not a formal strategic partnership or direct collaboration exists. This includes (1) hardware providers (e.g., manufacturers of the primary compute infrastructure), (2) infrastructure platforms (e.g., cloud service providers or data center operators that host the deployment), and (3) model sources

		(i.e., creators of foundational models used as the base for the project) ³
Sources	Official URL	Government publications and credible third-party reporting
Certainty	Certainty grade	See section on “Certainty ratings.” (Grade used for internal evaluation purposes; not visualized on webpage)
Defensibility	Defensibility score	See section on “Defensibility ratings.” (Score used for internal evaluation purposes; not visualized on webpage)
Sovereign Profile	Authoritative sovereignty framing for the project	Provides an attributed quote (with source URL) that states the sovereignty interest at stake

7) Sources and Verification

A. Primary sources (preferred)

- Official government publications and ministry announcements
- National AI strategy documents
- Budget documents and appropriations
- Government press releases
- Corporate documents relating to projects

B. Secondary sources (for validation)

³ Partner taxonomy: Partners are labeled by functional role. Common categories include: (1) technology providers—commercial firms supplying foundational hardware (GPUs/high-performance computing), cloud stacks, or base models used as inputs—(2) scientific/academic leads—universities, research institutes, and national research centers responsible for model architecture, data curation, or scientific direction—(3) operational partners—entities managing physical hosting, data center operations, system integration, or day-to-day execution—and (4) strategic/financial partners—entities providing capital, grants, cofunding, or supranational governance (e.g., EuroHPC Joint Undertaking, sovereign wealth funds). Rows may use more specific functional labels (e.g., “hardware provider,” “model source,” “cloud provider,” “system integrator,” “data contributor,” “host,” “cofunder”) that map to these categories.

- Media reports from credible and trustworthy media sources
- Reports from credible and trustworthy think tanks and nongovernmental organizations

C. Source quality standards

- All entries include a primary source URL
 - Funding figures are quoted verbatim from official sources
 - Entries are cross-referenced with multiple sources when details are ambiguous
 - official sources are prioritized over media reports
-

8) Data Quality Ratings

Ratings are used solely for internal evaluation purposes and are not displayed on the webpage. Projects are excluded from the index when either their certainty rating is “C” (low certainty) or their defensibility rating is “Needs review.” In all other cases, the research team makes a case-by-case determination, documented internally but not reflected in the published dataset.

Certainty ratings: Source quality and verifiability determine whether projects are included in the index. Projects are assigned one of three certainty grades:

- i)** A (high certainty): Multiple official government sources confirm key details (budget, partners, timeline). Information is current and verifiable. Projects included in the index
- ii)** B (medium certainty): Single official source or multiple secondary sources from credible outlets. Some details may require inference. Inclusion is determined by the research team through case-by-case assessment of source strength and corroborating evidence
- iii)** C (low certainty): Limited sources, relies primarily on media reports, or significant details are missing. Requires further verification

Defensibility ratings. Defensibility grades reflect how well a project satisfies the index's three core inclusion criteria—material government support, sovereign intent, and substantive AI deployment—and inform the research team's final inclusion decisions alongside certainty ratings:

- iv)** High: Clearly meets all three core requirements (material government support, sovereign intent, and substantive AI deployment)
 - v)** Medium: Meets core requirements but with ambiguities (e.g., unclear funding levels, indirect government involvement, or limited documentation)
 - vi)** Needs review: Borderline case that may not fully meet inclusion criteria. Requires additional research
-

9) Update Protocol

1. The index will be updated quarterly.
2. Events triggering changes to the index in the quarterly update:
 - a. New national AI strategy announcements
 - b. Major funding allocations or budget changes
 - c. Infrastructure procurement or deployment milestones
 - d. Status changes (e.g., Announced → Procurement → Buildout → Operational)
 - e. Project cancellations or significant scope changes
3. Sources monitored by research team:
 - a. Public lists and rankings such as TOP500 and EpochAI as independent and standardized reference points to corroborate claimed compute and documented model activity
 - b. Hugging Face as the main distribution platform for open-weight and open-source models/datasets

- c. Government publications such as ministry announcements, strategy documents, and budget releases
- d. Corporate and academic sources, including corporate blogs, ArXiv, and Ace
- e. Credible media reporting, including coverage from established technology and policy outlets
- f. Web searches, consisting of systematic keyword monitoring for new announcements and developments

10) Limitations and Considerations

1. Known Limitations

- a. **Funding comparability:** Funding figures may not be directly comparable across countries due to different accounting methods, timeframes, and purchasing power parity variations.
- b. **Project uncertainty:** “Announced” status projects may not materialize or may face significant delays.
- c. **Private investment opacity:** Private sector investment figures are often estimates and may not be fully disclosed.
- d. **Language and transparency barriers:** Some countries may have initiatives that have not yet been publicly announced or documented in English-language or readily translatable sources.

2. Classification Challenges

- a. Hybrid public-private projects require judgment by the research team on the degree of government involvement.
- b. Multipurpose high-performance computing (HPC) facilities upgraded for AI require an assessment of AI-specific investment.
- c. Regional collaborations require decisions on attribution (supranational vs. national level).

11) Contact & Contributions

This is a living document. Contributions, corrections, and updates are welcome. For additional questions or feedback about the methodology and inclusion criteria, please contact the research team at sovereignai@cnas.org.

Suggested citation (Chicago style):

Pablo Chavez, Vivek Chilukuri, and Ruby Scanlon, “CNAS Sovereign AI Index,” Center for a New American Security, April 20, 2026, <https://interactives.cnas.org/reports/sovereign-ai-index/>.

Last Updated: April 20, 2026

Next Scheduled Review: July 20, 2026