

Portuguese as a Programming Language in the AI Era

by Dinis Cruz and ChatGPT Deep Research, 2025/02/11

Context and Importance

A Global Language at a Crossroads: Portuguese is one of the world's most widely spoken languages, with roughly **260–270 million speakers** spanning Europe, South America, Africa, and Asia. It consistently ranks among the **top six languages globally by number of speakers** ([10 Fascinating Facts About the Portuguese Language](#)), and is the **third most spoken European-origin language** (after English and Spanish).

As the official language of **nine countries** and a working language in multiple international organizations, Portuguese wields significant cultural and geopolitical influence. In economic terms, the **Lusophone (Portuguese-speaking) world** collectively accounts for about **\$2.5 trillion in GDP** ([Community of Portuguese-Speaking Countries](#)(Source: [MOEA](#))) – roughly on par with the world's largest economies – underscoring its market potential and strategic importance.

Language in the Age of AI: Advances in *Large Language Models* (LLMs) and *Generative AI* (GenAI) have transformed human language itself into a powerful computing interface. Today, writing a prompt in natural language can execute complex tasks, blurring the line between coding and everyday speech. In effect, **language has become a new form of programming**. As NVIDIA CEO Jensen Huang recently observed, with modern AI “**the programming language is human**”, enabling anyone to instruct computers in plain English ([Nvidia CEO predicts the death of coding – Jensen Huang says AI ...](#)). This paradigm shift – sometimes summed up as “*English is the new programming language*” ([Thanks to AI, the Hottest New Programming Language is... English](#)) – carries profound implications: those languages best represented in AI will shape the future of technology and information flow.

Yet, *who* benefits from this revolution? So far, the lion's share of AI development has centered on English, creating an emergent asymmetry. Portuguese (along with many other languages) risks being left behind if AI systems are not taught to understand and “code” in it. The ability for Portuguese speakers to interact with AI in their **native language** is not just a matter of convenience – it is about **inclusive access** to technology, preservation of cultural-linguistic heritage, and equity in the digital economy. In this context, investing in Portuguese within AI is both a strategic opportunity and a necessity to ensure that the **LLM era** does not become a one-way street favoring English alone.

(Note: While this paper focuses on Portuguese as a case study, the insights and arguments apply to all European languages. Ensuring linguistic diversity in AI is a pan-European imperative.)

Language as the New Programming Paradigm

From Code to Conversation: Traditional software programming required learning specialized syntax and languages (Python, Java, C++, etc.). Today, with GPT-style LLMs, **any well-structured prompt in natural language can function as a program**, executing queries, generating content, or controlling devices. AI systems now interpret instructions, answer questions, and solve problems given in plain speech. This represents a seismic shift: **interaction in natural language is becoming the default user interface for computing tasks**. In practical terms, one can “write” an application by simply *telling* an AI what to do – a concept often described as *prompt engineering* or *zero-shot programming*. The rapid improvement of models (e.g. GPT-4, PaLM, Llama) has shown that carefully crafted English prompts can yield sophisticated outcomes that once required coding.

The English-First Bias: Current LLMs have largely been trained on massive English datasets, meaning English enjoys a first-mover advantage as the de facto *programming language* of AI. Complex prompt techniques, knowledge repositories, and even documentation for AI models are predominantly in English. As a result, an English-speaking user can leverage AI with greater precision and reliability today than a user issuing instructions in Portuguese (or Polish, or French). Research indicates that AI tools trained on internet data risk widening the gap between speakers of “data-rich” languages and others ([How language gaps constrain generative AI development](#)). In other words, if AI understands English best, English speakers reap disproportionate benefits, and global linguistic diversity may be subtly eroded ([How AI threatens linguistic diversity - Opinion - Chinadaily.com.cn](#)). This is not a mere academic concern: **language bias in AI** could translate into fewer services, less information, and diminished economic opportunities for those operating in languages outside the AI mainstream.

Portuguese as Code: The solution is clear – *equip Portuguese to stand alongside English as a programming medium for AI*. We must develop AI models that *think* and *operate* in Portuguese with the same fluency and capability that today's top models demonstrate in English. This goes beyond basic translation. It means training AI to accurately grasp Portuguese idioms, context, and nuance; to access and output knowledge stored in Portuguese; and to allow Portuguese instructions to accomplish anything an English instruction could. In essence, **LLMs should treat Portuguese as an equal “first-class” language for interfacing with technology**. When a Portuguese policymaker can query a national AI system in Portuguese and get a detailed policy brief, or a Brazilian entrepreneur can prototype an app through Portuguese prompts, we will have achieved parity. The tools are within reach – the latest generation of LLMs are inherently multilingual – but concerted effort is needed to optimize and localize them. Just as English became the initial programming language of AI by virtue of data dominance, **Portuguese (and other European languages) can become languages of AI by deliberate investment and strategic focus**.

The Case for Portuguese: Data and Impact

To understand why prioritizing Portuguese in AI makes sense, consider the following **key facts and figures** about Portuguese in the world today:

- **Massive Global Speaker Base:** Portuguese has approximately **265–270 million total speakers**, making it the **6th most spoken language worldwide** ([10 Fascinating Facts About the Portuguese Language](#)). It is the **official language in 9 countries** and spoken across **4 continents**, from Brazil (over 210 million speakers) to Angola, Mozambique, Portugal, and beyond. This widespread use means AI services in Portuguese have a vast addressable audience.
- **Strong Digital Presence:** Portuguese is a major language on the internet. As of 2020, there were over **171 million internet users** communicating in Portuguese, about **3.7% of the global internet population** ([Languages used on the Internet - Wikipedia](#)). By number of internet users, Portuguese ranks **#6 on the Internet** (just behind Hindi and French) ([How is the Presence of the Portuguese Language on the Internet ...](#)). Moreover, roughly **67% of Portuguese speakers are online**, a rate that continues to grow with expanding connectivity in Brazil and Lusophone Africa. This indicates a robust and engaged online community ready to leverage AI tools in Portuguese.
- **Economic Power of Lusophone Markets:** The combined GDP of Portuguese-speaking countries is roughly **\$2.4–2.5 trillion USD** ([Community of Portuguese-Speaking Countries](#)(Source: [MOEA](#))). Brazil's economy (the world's 9th largest) and Portugal's EU membership make Portuguese economically significant. Lusophone Africa (Angola, Mozambique, etc.) adds emerging markets with young, growing populations. An AI

ecosystem that serves Portuguese effectively can tap into these economies, driving innovation in sectors from e-commerce and fintech to education and healthcare.

- **Cultural and Linguistic Influence:** Portuguese is a **language of science, literature, and diplomacy**. It is the language of Camões and Saramago, of Bossa Nova and Morna, of cutting-edge research published in Brazilian journals. It is also one of the fastest-growing European languages; for instance, Brazil's population and influence ensure Portuguese remains a top world language for generations. By embedding Portuguese in AI systems, we also ensure that the rich knowledge encoded in Portuguese texts (archives, libraries, Wikipedia, news media) is not lost to the AI revolution. Importantly, we preserve cultural nuances – AI that understands Portuguese proverbs, legal terms, or medical vocabulary can provide more relevant and accurate responses for Portuguese speakers.

Impact of Investment: Investing in AI for Portuguese means unlocking this potential. It means an Angolan student can use a tutoring chatbot that understands *her* language and context; a Brazilian doctor can consult a medical AI trained on Portuguese case studies; a Portuguese journalist can quickly search vast Portuguese archives via an AI assistant. The social impact (greater inclusion, literacy, services access) and economic upside (new markets, improved efficiency, local AI industry growth) are substantial. In short, **the data paints a clear picture: Portuguese is too globally important to ignore in the AI era**. Any nation or business strategy that values scale, inclusivity, and long-term relevance should treat Portuguese-language AI competence as a strategic priority.

Strategic Actions to Prioritize Portuguese in AI

Ensuring that Portuguese thrives as a “programming language” of AI will require coordinated effort across government, industry, and academia. Below are key strategic actions and initiatives that stakeholders should undertake:

1. Invest in Native AI Models and NLP Tools for Portuguese

Build AI by and for Portuguese speakers. The foundation of language-enabled AI is data and models. Governments and enterprises should **invest in creating large-scale Portuguese language models** – from foundational LLMs to specialized tools – rather than relying solely on translations of English-centric systems. Notably, multilingual models exist, but dedicated Portuguese models can achieve higher fidelity. For example, the *BERTimbau* project produced a state-of-the-art Portuguese NLP model that **outperformed multilingual BERT on Portuguese tasks** ((PDF) [BERTimbau: Pretrained](#)

[BERT Models for Brazilian Portuguese](#)), underscoring the value of native-language AI research. More recently, a consortium of Portuguese researchers announced **AMÁLIA**, the first large-scale **Portuguese-only LLM**, with a planned release by 2026 ([Final version of Portuguese large language model launched in 2026](#)). Such efforts should be accelerated and expanded. Key investments include:

- **Training Data and Compute:** Curate vast Portuguese text datasets (e.g. web archives, literature, official documents) to train models. Leverage national computing infrastructure or cloud partnerships to train high-parameter models comparable to GPT-3/4 – but in Portuguese. Initiatives might involve creating a **Portuguese GPT** or multilingual GPT with enhanced Portuguese capability.
- **NLP Tools and Resources:** Support the development of **natural language processing (NLP) tools** such as tokenizers, speech recognition, and text-to-speech systems tailored to Portuguese. For instance, accurate Portuguese speech assistants and transcription services can make AI more accessible.
- **Benchmarking and Optimization:** Create **benchmark datasets** and challenges (for question answering, summarization, dialogue in Portuguese) to track progress and encourage competition. Optimize models for not just European Portuguese but also Brazilian Portuguese and regional dialectal differences, ensuring inclusivity within the Lusophone world.

Investing in homegrown models not only improves performance for Portuguese users, but also secures **technological sovereignty** – reducing dependence on foreign AI providers and ensuring local control over data and ethics. It sets the stage for Portugal (and Lusophone partners) to export AI solutions to other language markets, turning a linguistic asset into a competitive advantage.

2. Foster Open-Source Collaboration and Government AI Initiatives

Leverage the power of community and public support. Open collaboration has proven immensely successful in AI – the best example being **BLOOM**, a 176-billion-parameter open model that can generate text in 46 languages (including Portuguese) and was produced by a global volunteer effort ([BLOOM - BigScience](#)). Portugal and other Portuguese-speaking countries should actively participate in and initiate **open-source AI projects** centered on language. This includes contributing to international projects and launching dedicated programs for Portuguese.

Governments play a catalytic role here. **Policy makers should back open research and infrastructure** for AI in Portuguese. This could mean funding a **“Portuguese AI Commons”** – open datasets (e.g. the BrWaC web corpus with billions of words), open models, and APIs that startups and researchers

can use. It also means adopting favorable policies: for example, requiring that publicly funded AI research be open-source, and encouraging data sharing agreements (while respecting privacy) so that Portuguese language data from libraries, media, and academia can fuel AI development.

There are positive precedents to build on. The European Union's Digital Europe Programme emphasizes **multilingual AI technologies** and could support Portuguese projects ([eLangTech: The EU's Multilingual toolset - Nimdzi](#)). The European Language Equality roadmap aims for **full digital language equality by 2030** ([European Language Equality](#)) – a vision that aligns perfectly with boosting Portuguese in AI. Portugal's national AI strategy ("AI Portugal 2030") can integrate language objectives, ensuring that resources are allocated to NLP and that Portuguese is well-represented in European AI initiatives. Collaboration should also extend to the **Community of Portuguese Language Countries (CPLP)** – a united effort among Lusophone nations to share data, talent, and applications (for example, building a common Portuguese AI translation system or a joint research center). By pooling efforts in an open, transparent manner, the Portuguese-speaking world can punch above its weight in the AI arena, producing tools that benefit all and are freely available. Open-source also invites global talent: researchers from anywhere can contribute to Portuguese AI, and Portuguese contributions to multilingual projects will raise the language's profile in the AI community.

3. Build AI-Driven Knowledge Graphs and Semantic Systems

Connect the Portuguese knowledge universe. A language is not just grammar and vocabulary – it's a repository of knowledge. One strategic move is to construct comprehensive **knowledge graphs** and **semantic databases** for Portuguese information. These are structured networks of facts and concepts (people, places, events, terms) and their relationships, which AI can query to retrieve reliable answers. Imagine an AI that can answer a question like "What were the economic effects of the 1755 Lisbon earthquake?" by consulting a Portuguese knowledge graph that links historical records, economic data, and scholarly research – all in Portuguese. This is possible if we invest in the creation of such semantic infrastructure.

Projects could include:

- **Portuguese Knowledge Graphs:** Using Wikipedia in Portuguese, local databases, and academic content to build a graph of interconnected entities in the Lusophone context. This would enable advanced **question-answering in Portuguese**, as demonstrated by research on using knowledge graphs for QA ([\[PDF\] Can SPARQL Talk in Portuguese? Answering Questions in Natural ...](#)). Users could query in natural Portuguese and get precise answers grounded in verified data (crucial for applications like fact-checking ([\(PDF\) Fact-Checking for Portuguese: Knowledge Graph and Google ...](#)), education, and expert systems).

- **Multilingual Semantic Integration:** Ensure that Portuguese knowledge systems interface with global ones. For instance, align the Portuguese graph with DBpedia/Wikidata entries so that AI can seamlessly transition between languages when needed, and so that content available only in Portuguese is not invisible to the wider AI ecosystem.
- **Domain-Specific Ontologies:** Develop semantic ontologies in critical domains – e.g. medicine, law, energy – in Portuguese. This would allow specialized AI assistants (a legal advisor AI, a medical diagnostic AI) to understand the precise meaning of Portuguese legal codes or medical terminology and reason over data.

By building these semantic assets, we **augment the “brain” of AI with Portuguese context**. It ensures that AI systems not only speak Portuguese, but truly **understand the world through Portuguese**. This is key for governmental use (e.g. policy intelligence systems that parse Portuguese policy documents) and for businesses (e.g. semantic search in Portuguese for enterprise data). Such knowledge graphs can be built through partnerships between universities, libraries, and tech companies, supported by government grants. Over time, this creates a **virtuous cycle**: as AI uses the knowledge graph, it can also help expand and refine it, leading to continuously improving Portuguese knowledge accessibility.

4. Support Portuguese-Based AI Startups and Innovation Hubs

Cultivate an ecosystem of innovation. The entrepreneurial community will be instrumental in turning language AI capabilities into real-world applications. We need to empower **startups and tech companies focusing on AI solutions for the Portuguese-speaking market**. This can be achieved through a mix of incentives, funding, and incubator programs:

- **Funding and Incentives:** Governments and investors should allocate dedicated funding for **Portuguese-centric AI startups**. For instance, venture funds or grant programs can target NLP, speech tech, or AI services that cater to Portuguese users. Tax breaks or innovation vouchers could encourage companies to develop Portuguese-language features in their products. Recent trends are encouraging – Portugal’s startup ecosystem attracted over **\$1 billion in funding in 2021** (doubling from the previous year) ([Discover Portugal: An Unbeatable Value Proposition for Startups](#)), and AI is a growing area within this surge. By directing some of this investment towards language-focused AI, stakeholders can ensure this growth benefits Portuguese users directly.
- **Innovation Hubs and Clusters:** Establish **AI innovation hubs** in Lisbon, São Paulo, Luanda, and other Lusophone tech centers, where startups, researchers, and industry can collaborate. These hubs can host accelerator programs (for example, a bootcamp for AI in education technology

with Portuguese as the medium), provide mentorship, and facilitate access to resources like cloud computing or datasets. Portugal's hosting of major tech events (such as the Web Summit) and the presence of research institutes can be leveraged to attract global attention to Portuguese AI innovation.

- **Public-Private Partnerships:** Encourage partnerships where large enterprises (e.g. banks, telecommunications firms, healthcare providers) pilot new AI solutions in Portuguese developed by startups. This not only gives startups a testbed and revenue, but also drives adoption of Portuguese AI in important sectors. For example, a collaboration could develop a customer service AI that handles queries in Portuguese across various dialects, improving service for millions of customers while advancing NLP capabilities.

By supporting startups, we also **create local expertise and high-value jobs**. Rather than a brain drain where top AI talent leaves for Silicon Valley, a vibrant local market will keep Portuguese AI engineers and researchers engaged at home, or even attract foreign talent interested in multilingual AI. Over time, a successful cohort of AI companies focusing on Portuguese can expand globally – exporting their tech to other language markets – thus turning an initial focus on linguistic inclusion into a competitive export advantage.

In summary, these strategic actions – investment in models, open collaboration, semantic systems, and startup support – form a comprehensive approach. They attack the challenge from all sides: technology, data, knowledge, and market deployment. The coordinated execution of these steps by government agencies (through funding and policy), by academic institutions (through research and training of talent), and by the private sector (through innovation and scaling) will establish Portuguese as a **first-class citizen in the AI universe**.

Policy and Industry Recommendations

To operationalize the above strategy, we present targeted recommendations for different stakeholder groups. These concrete proposals aim to integrate Portuguese deeply into AI frameworks and ensure sustained support:

- **For Government & Policymakers:**
 - **Integrate Language in AI Policy:** Make support for the Portuguese language a pillar of national AI strategies (e.g., *AI Portugal 2030*). Set explicit goals for Portuguese AI capability (such as top-tier translation quality, or a domestic Portuguese GPT by a certain date) and align public R&D programs to achieve them.

- **Public Investment in R&D:** Increase funding for NLP and language tech through national science foundations and innovation agencies. Sponsor large-scale projects like corpus creation, or the **AMÁLIA Portuguese LLM initiative**, to ensure they reach fruition. Consider establishing a **Lusophone AI Institute** that coordinates research across Portuguese-speaking countries.
- **Open Data and Infrastructure:** Implement policies that open up government data (parliament transcripts, public libraries, educational content) in Portuguese for AI use. Create a national repository for Portuguese datasets and provide compute resources (perhaps via a state-funded supercomputing center) for researchers and companies training models.
- **Education and Skills:** Incorporate AI and language technology in education curricula. Train the next generation of AI professionals with a multilingual mindset – for example, encourage NLP courses to include assignments on Portuguese data. Support scholarships and exchanges focused on computational linguistics for Portuguese.
- **For Businesses & Industry Leaders:**
 - **Prioritize Language Inclusion:** When developing AI products (be it a virtual assistant, a search engine, or an analytics tool), ensure Portuguese language support is not an afterthought but a core feature. The Portuguese-speaking market is huge – catering to it can be a major competitive edge.
 - **Invest in Startups:** Corporate venture arms and local investors should look to **Portuguese-focused AI startups** for investment. These startups not only address a niche that global players might overlook, but also can scale to other languages once proven. Your backing can accelerate solutions in areas like Portuguese speech interfaces, local dialect chatbots, or AI for Portuguese education.
 - **Corporate-Government Collaboration:** Partner with public initiatives (such as a government project to deploy AI in public services in Portuguese). Businesses can offer expertise and in-kind resources, and in return shape the solutions and benefit from public sector adoption. For example, a consortium of tech companies could collaborate on a **Portuguese AI assistant for e-government services**, making interactions with citizens more efficient while demonstrating their technology.
 - **Join Open Initiatives:** Encourage your developers and data scientists to contribute to open-source projects (like adding Portuguese support to international AI frameworks, or contributing code to Portuguese NLP libraries). Not only does this build goodwill, it also ensures your company stays at the cutting edge of language AI developments.
- **For AI Developers & Researchers:**

- **Embed Multilingualism:** When building models or datasets, include Portuguese data and test cases. Even if your primary target is not Portuguese, adopting a multilingual approach (as done with models like BLOOM) improves robustness and broadens impact. Push for Portuguese versions of important benchmarks and publish results on them.
- **Create and Share Tools:** Develop open-source tools that ease working with Portuguese – such as tokenizers, stemming libraries, or prompt repositories tailored to Portuguese. Sharing these lowers the entry barrier for others to do Portuguese AI work. For instance, sharing a well-crafted prompt that solves a task in Portuguese can save others time and encourage more experimentation in that language.
- **Research Language-Specific Challenges:** There are interesting NLP research questions in Portuguese (e.g., handling verb conjugation complexity, or disambiguating formal vs. informal address). Tackling these can lead to publications and also improve technology. Seek out these challenges and collaborate with linguists and domain experts.
- **Community and Events:** Participate in or organize workshops, hackathons, and conferences on **Portuguese NLP and AI**. Having dedicated events (potentially under larger AI conferences or within Lusophone countries) raises the profile of the language. A “Portuguese NLP Summit” or inclusion of Portuguese tasks in international competitions (like a Portuguese question-answering challenge in a global contest) would galvanize interest.

By following these recommendations, each stakeholder contributes to a holistic ecosystem where Portuguese is thoroughly integrated into AI development cycles. The government ensures resources and a favorable environment; industry drives application and scale; and the tech community pushes the boundaries of what’s possible. This multi-pronged collaboration is crucial – no single actor can achieve language parity in AI alone, but together it’s attainable.

Conclusion and Call to Action

A Future Worth Building: Portuguese is at a pivotal moment in the AI and LLM era. The choices made now will determine whether it **flourishes as a digitally empowered language** or gets sidelined in favor of more dominant tongues. The evidence and arguments presented here make a compelling case that investing in Portuguese for AI is both strategically wise and morally sound. It is an investment in **people** – the hundreds of millions who speak Portuguese and deserve technology that understands them. It is an investment in **innovation** – unlocking new solutions and markets by

leveraging a great world language. And it is an investment in **diversity** – reinforcing a future where AI reflects the rich tapestry of human language and culture, rather than homogenizing it.

The call to action is clear: **policy makers, entrepreneurs, researchers, and media influencers** must recognize language as the backbone of AI innovation. We urge governments in Portuguese-speaking nations and the EU to **treat language equality as a tech priority**, committing to the necessary funding and frameworks. We urge companies and startups to **seize the Portuguese opportunity** – those who do so will not only access a large market but also set themselves apart as leaders in a less crowded space. We urge the AI community to **champion multilingual and open approaches**, so that breakthroughs in AI benefit all languages, not just a few.

The era of “English-only” AI is already evolving. In the coming years, as projects like Europe’s language equality roadmap drive progress ([European Language Equality](#)), we can expect a landscape where interacting with AI in **Portuguese (or Spanish, French, German, etc.) is just as seamless as in English**. Achieving this will require dedication and collaboration, but the reward is enduring: a world in which technology empowers individuals *in their own language*, and where Portuguese, with its history, vitality, and global reach, stands as a **fully enabled programming language of intelligent machines**.

Now is the time to act and shape that future. **Portuguese has given poetry, knowledge, and connection to the world for centuries; let us now give it a prominent place in the AI revolution**. By doing so, we not only honor the linguistic heritage of millions but also unleash the full potential of AI to serve humanity in all its diversity. This is a call to **invest, innovate, and include** – so that the next chapter of the digital age is written *em português*.