

In *Público* newspaper, February 9, 2023, p.9

https://www.publico.pt/2023/02/09/opiniao/opiniao/lingua-portuguesa-tecnologia-futuro-2038078?ref=opiniao&cx=page_section

Translated into English by the authors

Portuguese language and technology: the future is now

The advances of Artificial Intelligence have been impressive, especially in its application to Language Technology. This progress is based on machine learning with so-called Large Language Models, such as GPT-3 or ChatGPT among others, which have been talked about so much recently.

These models are artificial neuronal networks that learn by adjusting the weights of the connections between neurons to improve their accuracy level during training. Training tasks consist of predicting omitted words from their contexts of occurrence.

These networks are gigantic — GPT-3, for example, has 175 billion connections between neurons. They capture linguistic regularities when trained in massive computational processes, on colossal volumes of language data, text or audio. As for GPT-3, 500 billion words were used in its training, whose computing cost is estimated at more than \$4 million.

Once trained, these models can be used in other language processing tasks with an unprecedented level of quality, such as translation, conversation, speech transcription and subtitling, text and speech generation, content analysis and information extraction, etc. When integrated into larger systems, they are transforming diagnostics and health care, financial and legal services, gaming and entertainment, education, creativity and culture, etc.

Due to the size of the respective models, these processing tasks are available remotely as online services, as it is the case with search engines, and not as the spelling checkers, locally installed in our devices. Due to the size of the resources for machine learning, in the immediate future these services are made available by the oligopoly of bigtechs, which you can count on the fingers of one hand, with the capacity to access the colossal volumes of computing and data needed for training.

As a result, in the digital age the use of language — with other human beings, organizations, services or artificial devices — will not be made again without this pervasive and profound technological intermediation, which processes communication acts and accesses its meaning.

We have enough experience with search engines, for example, and with their assumptions and impacts, to anticipate the consequences of this technological intermediation in the daily use of language itself. Technological intermediation, in general, generates a digital trail of personal data outside our control. Unceasing technological intermediation of human language and communication, funneled into a small global oligopoly, creates alarming risks to individual and collective sovereignties.

Undesirable impacts of emerging technologies are mitigated with more and better technology, not with less. The dispersion of the provision of these services is crucial to deter the threat posed by their concentration. The answer relies thus in the promotion of an ecosystem of innovation that, alternatively, timely allows trivial access to the resources necessary for the appropriation and exploitation of Language Technology by the largest possible number of individuals and organizations, private and public, small and large, national and international.

Given the volumes of data and computing costs that need to converge and be made available to leverage such an ecosystem, and in the face of the most relevant public interest at stake, this is a new and urgent task for democratic states, alone and in cooperation.

The Portuguese language, with 250 million speakers on 4 continents, is one of the largest international languages of global projection. The indicators point to its growth until the end of the century, with the majority of speakers in the African continent. However, if to classical language promotion policies, we do not add a clear promotion of its technological preparation, it will lose importance and tend to be replaced by other languages. For this reason, it is important to join forces for a Technological Preparation Plan for the Portuguese Language, which will have effects in a vast range of sectors of activity, including the whole economy at large.

In Spain, for more than a decade, the Language Technologies Leverage Plan has been deployed: between 2015 and 2020, it received an allocation of EUR 89 million, reinforced with an additional allocation of EUR 334 million until 2025, now under the New Language Economy initiative. Several European initiatives can be used - if we believe that the Portuguese language is an asset that we cannot afford to waste.

Ana Paula Laborinho
Director in Portugal
OEI Organization of the Ibero-American States

António Branco
General Director
PORTULAN CLARIN Research Infrastructure for the Science and Technology of Language