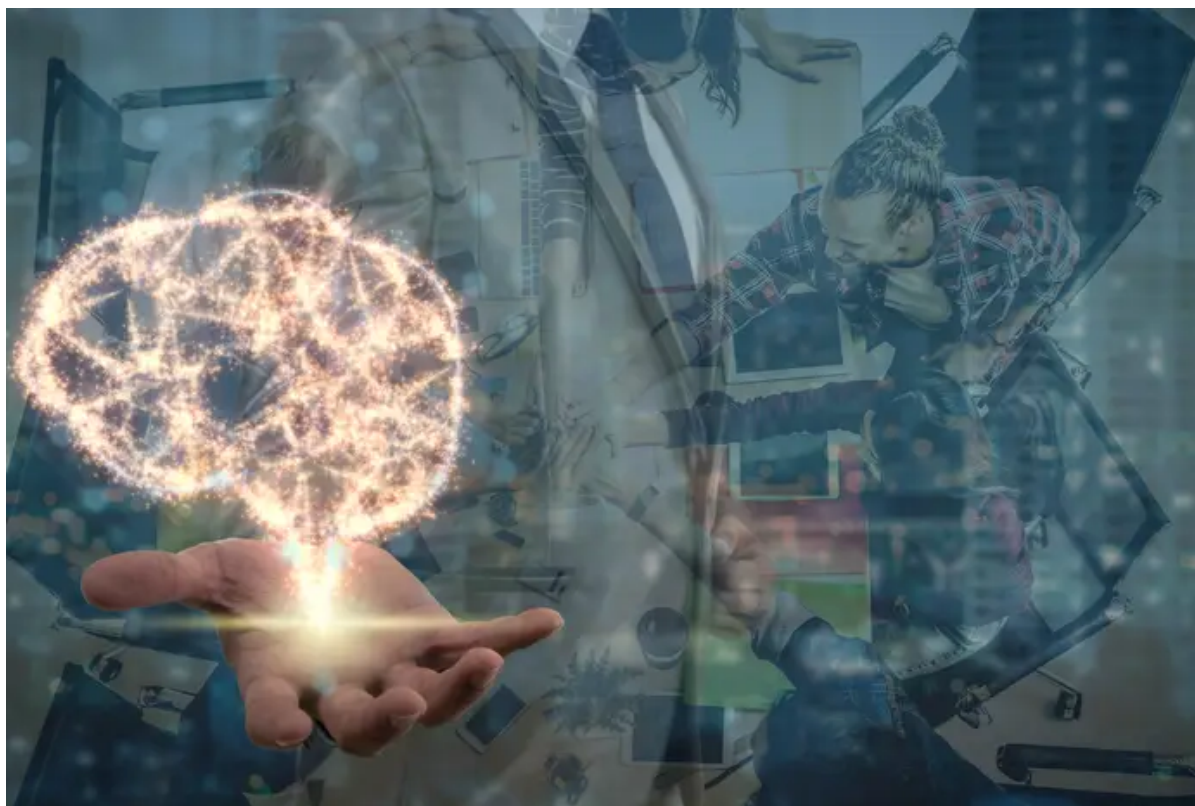




**Exclusivo**

**TECNOLOGIA**

# Lançada a primeira Inteligência Artificial que gera textos em português sobre qualquer tema. Chama-se Albertina



Photographer is my life/Getty Images

O primeiro modelo de inteligência que pesquisa dados e produz textos em português arranca com um módulo de 900 milhões de parâmetros. Em julho segue-se um segundo modelo que “percebe” português nas variantes europeia e brasileira



15 MAIO 2023 19:40

**Hugo Séneca**

**A**lbertina é um nome com tradição, mas foi escolhido por investigadores da Faculdade de Ciências da Universidade de Lisboa (FCUL) e Faculdade de Engenharia da Universidade do Porto (FEUP) também para apontar para o futuro com o primeiro modelo de Inteligência Generativa, que permite produzir textos com vários temas, de forma completamente autónoma tanto nas variantes europeia e brasileira, da língua portuguesa. Desde a discreta estreia, na sexta-feira, até esta segunda-feira, a Albertina já foi descarregada 33 vezes – um número que não é assim tão reduzido se se tiver em conta que esta ferramenta foi desenhada para correr em servidores (computadores empresariais, de grande capacidade).

Este é apenas o primeiro módulo. Em julho, o projeto deverá ficar concluído com um segundo módulo conhecido por GPT-PT, que será grátis, à semelhança da Albertina. “A Albertina é o primeiro modelo de processamento de linguagem para a língua portuguesa que é [disponibilizado publicamente](#), tanto para atividades académicas como para comerciais, sem custos e com o código aberto (os produtores disponibilizam estrutura e códigos de programação). As grandes marcas tecnológicas também hão de ter qualquer coisa do género, porque já conseguem lidar com o português, mas não disponibilizam esses modelos livremente como nós estamos a disponibilizar”, explica António Branco, professor da FCUL.

O ChatGPT é provavelmente o caso mais famoso daquilo que se consegue fazer com os grandes modelos de processamento de linguagem (GPT), que usam inteligência artificial para gerar textos a partir de grandes repositórios de dados. A Albertina chega com a expectativa de abrir caminho a ferramentas similares ao ChatGPT, que foi criado a partir dos GPT produzidos pela OpenAI (GPT 3.5, ou mais recentemente GPT 4).

À semelhança do que se passa com as empresas que lideram o segmento,

a Albertina funciona como recurso tecnológico que serve de base ao desenvolvimento de várias ferramentas, que produzem conteúdos úteis para o comum dos utilizadores: “A Albertina pode ser usada para produzir sumários, traduções, classificação de texto consoante as emoções, diálogos com pessoas (e assistentes digitais), legendagem automática, entre outras coisas”, acrescenta António Branco.

Os 33 downloads efetuados desde sexta-feira levam a crer que, tanto para efeitos académicos como para efeitos comerciais, a Albertina já começou a ser explorada para o desenvolvimento de novos produtos e tecnologias. António Branco confirma que, até à data, existia na variante brasileira um protótipo bem “mais pequeno” que a Albertina, e lembra que este é o primeiro GPT em português europeu.

O professor da FCUL considera o lançamento da Albertina um marco histórico com significado comparável ao lançamento da primeira gramática para o português. E por isso a iniciativa é também encarada como um reforço da soberania linguística.

O projeto foi desenvolvido no âmbito do consórcio Accelerate.AI, que recebeu 35 milhões de euros para o desenvolvimento de diferentes projetos, ao abrigo do Programa de Recuperação e Resiliência (PRR). Na Inteligência Artificial Generativa, o Accelerate.AI prevê o desenvolvimento de um codificador e de um decodificador. O codificador é a Albertina. O decodificador é o já mencionado GPT-PT que há-de sair em julho pelas mãos dos investigadores da FCUL.

Ambos podem fazer as mesmas coisas, mas enquanto a Albertina requer treinos mais intensivos e especializados em diferentes temáticas, o GPT-PT tem apenas um treino mais básico que lhe permite adaptar-se mais facilmente a qualquer temática para produzir textos de forma autónoma, com margens de erro menores.

A Albertina exige mais treino por estar assente numa rede neuronal de menor dimensão. Em contrapartida, o GPT-PT tem uma rede neuronal maior e mais complexa e, por isso, aprende mais rapidamente a produzir dados.

Há um traço em comum entre codificador e decodificador: ambos produzem resultados sem que os humanos tenham de inserir regras de processamento de dados e produção de conteúdos, recorda António Branco.



António Branco, professor da Faculdade de Ciências da Universidade de Lisboa, recorda que a Albertina é um marco histórico comparável ao lançamento da primeira gramática portuguesa **DR**

A Albertina está assente numa rede neuronal artificial que, como o nome indica, pretende funcionar da mesma forma como os neurónios se interligam. Dentro de um cérebro, essas interligações são feitas por sinapses, mas num GPT são conhecidas por “parâmetros”. Estes parâmetros mais não são do que valores estatísticos que definem como é que um GPT vai passar a interpretar o mundo – ou pelo menos, quais as abordagens que vai seguir para o processamento da informação.

No caso de codificadores como a Albertina, o treino pode ser feito ao longo do tempo e variar consoante a temática. Num decodificador como o GPT-PT o treino resume-se à fase de preparação – e não há que voltar aos treino durante a fase de operacionalidade.

Em qualquer destes dois casos, o conceito tem por ponto de partida a introdução de textos que são processados a partir da identificação de padrões estatísticos que ajudam a compreender não só significados, como também a sintaxe, contextos ou episódios históricos. É com base nesta capacidade que os GPT produzem textos tentando adivinhar, através das probabilidades estatísticas, quais as palavras que faltam para compor uma frase - ou um texto inteiro sobre um tema à escolha.

António Branco admite que a livre distribuição destas ferramentas pode ser aproveitada por cibercrime ou brincadeiras de mau gosto, mas garante que está a seguir as melhores práticas: “O que estamos a fazer com esta distribuição gratuita da Albertina também já foi feito para inglês, espanhol ou francês”, sublinha.

O professor recorda que os codificadores de língua inglesa têm um número de parâmetros similar aos 900 milhões da Albertina, mas também admite que nos descodificadores, em que se há-de enquadrar o futuro GPT-PT, o número de parâmetros é bem maior, como sucedeu com o GPT 3.5 (175 mil milhões) que está na origem dos primeiros casos de sucesso do ChatGPT, da OpenAI.

“Já foi lançado o GPT 4 pela OpenAI, mas nunca foi publicado um artigo a explicar como funciona e ninguém conhece as características técnicas”, recorda António Branco, dando mais um argumento para a defesa da soberania linguística face aos projetos dominados pelas grandes marcas de tecnologias.

A Albertina foi treinada com textos redigidos tanto em português europeu como português do Brasil. Em qualquer dos casos, a eficiência na busca e no processamento de informação depende da quantidade de dados disponíveis. “Admito que não vamos conseguir levar o GPT-PT a ter a mesma dimensão dos GPT da OpenAI. E não é por uma questão técnica ou de falta de recursos, mas sim por não haver, na Internet, tantos textos disponíveis em português, como há atualmente para inglês”, conclui o investigador da FCUL. Em julho se verá o que traz o GPT-PT.