

Title: Jais (language model)

URL: [https://en.wikipedia.org/wiki/Jais_\(language_model\)](https://en.wikipedia.org/wiki/Jais_(language_model))

PageID: 75006237

Categories: Category:2023 in artificial intelligence, Category:2023 software, Category:Arabic language, Category:Generative pre-trained transformers, Category:Large language models, Category:Natural language processing, Category:Open-source artificial intelligence, Category:Science and technology in the United Arab Emirates

Source: Wikipedia (CC BY-SA 4.0).

Jais is an open-source large language model launched in August 2023. Developed as a collaboration between Emirati AI company G42 , the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), and US-based Cerebras Systems , Jais was designed to produce high-quality Arabic text and was also trained on English data. [1] [2]

The model's creation was motivated by the underrepresentation of the Arabic language in the field of generative artificial intelligence . It aims to provide a more culturally and linguistically accurate model for the world's 400 million Arabic speakers. [3] Its name is a reference to Jebel Jais , the highest mountain in the UAE. [2]

Background and development

Jais was developed in response to the limited availability of advanced generative artificial intelligence models for the Arabic language, despite it being spoken by over 400 million people. [3] Existing models were often trained on limited or low-quality Arabic web content, resulting in poor performance. [4] The project represents a significant investment by the United Arab Emirates in the field of AI as part of its national strategy. [1]

The model was created through a partnership between Inception (now Core42), a subsidiary of the Abu Dhabi -based AI company G42 ; the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI); and Cerebras Systems , a US company specializing in AI hardware. [2] [1] The model is named after Jebel Jais , the highest peak in the UAE. [2]

Training

The initial version of Jais released in August 2023 had 13 billion parameters . In November 2023, Core42 released Jais 30B , an improved version with 30 billion parameters. [5] Both models were trained on a subset of the Cerebras Condor Galaxy 1 supercomputer. [2] [1]

The training dataset consisted of a mix of Arabic, English, and computer code. [2] [3] According to Timothy Baldwin, a professor of natural language processing at MBZUAI, training the model on a diverse Arabic dataset allows it to switch between dialects. [3]

Features

Jais is designed to generate text in both English and Arabic. The project has also released instruction-tuned "Chat" variants for both the 13B and 30B models, which are specifically optimized for conversational applications. [5] Additional functionality for working with images, graphs, and tabular data is planned for future releases. [3]

References

External links

Official website