



Yogesh Haribhau Kulkarni • You

AI Advisor (Helping people/organizations in their AI jo...
now • 🌐

Aya, simply put, is a worldwide collaborative initiative to prepare language datasets for building a new massively multilingual, generative LLM that outperforms existing open-source models and covers 101 different languages.

Thrilled to share that the model and the dataset are now available for public use. <https://lnkd.in/dENTahUG> This is a significant step in accelerating multilingual AI. Cohere For AI

I actively contributed to this project over the past year, making 5k+ dataset contributions. Delighted to be recognized as one of the top 50 contributors globally and proud to hold the number 1 spot for Marathi, my mother tongue.

Some of the salient points about AYA are:

➡ Aya is intentionally fully multilingual first – Most generative AI models are designed to be English first, and many global communities have been left unsupported due to the language limitations of existing models.

➡ The Aya model is state-of-the-art and outperforms existing massively multilingual open-source models – The Aya model covers 101 different languages, far ahead (2x) of open-source baselines (mt0, bloomz). It outperforms the best multilingual models in benchmark tests while extending language coverage.

➡ In addition to the Aya model, we are releasing the Aya collection – the largest collection of multilingual instruction fine tuned dataset to date, with 513 million prompts and completions covering 114 languages. We fully open source the

collection, which includes rare human curated annotations from fluent speakers worldwide.

➡ Largest participatory research initiative to date, changing how breakthroughs happen – The research involved 3K independent collaborators across 119 countries, making this the largest open science project to date in the field of machine learning.

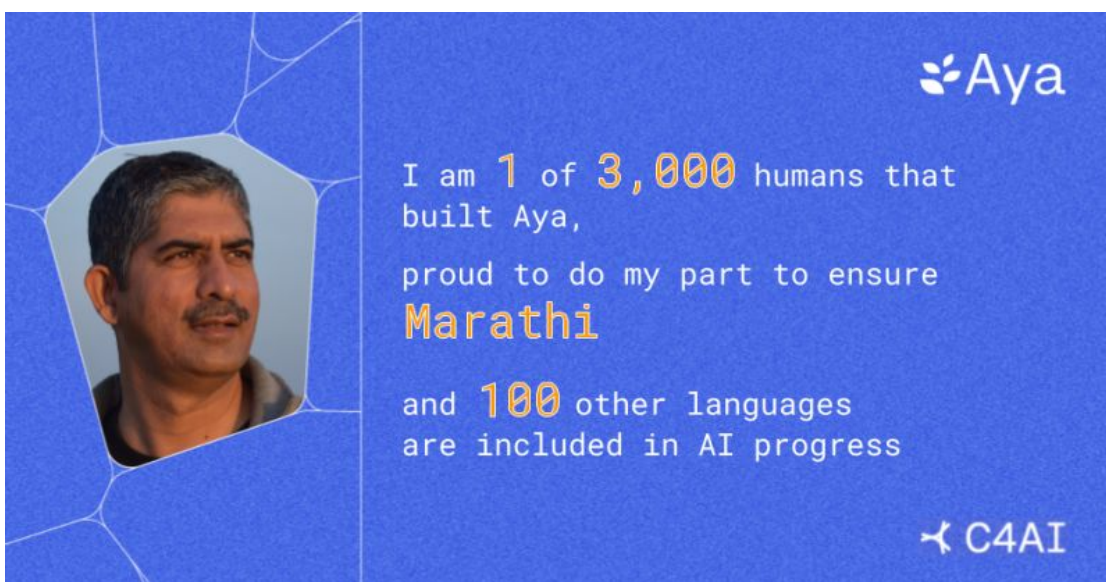
➡ C4AI is open-sourcing this work to unlock AI research on a global scale – Aya will be made available to the research community to further advance this critical effort. The work provides AI researchers with a first-of-its-kind foundation to support multilingual AI research projects.

#LanguageModels #aya #language #India #Marathi
#AyaProject #MultilingualAI #ai #ml #nlp #generativeai #llms
#dataset #opensource #gde #mvp #mvpbuzz

Sara Hooker Madeline Smith Shivalika Singh Abinaya
Mahendiran

Krutrim Ola Bharat GPT BHASHINI Sarvam

AI KissanAI Cohere Google Microsoft Google Developer
Experts TensorFlow User Group (TFUG) GDG Pune GDG Cloud
Pune



Like

Comment

Repost

Send