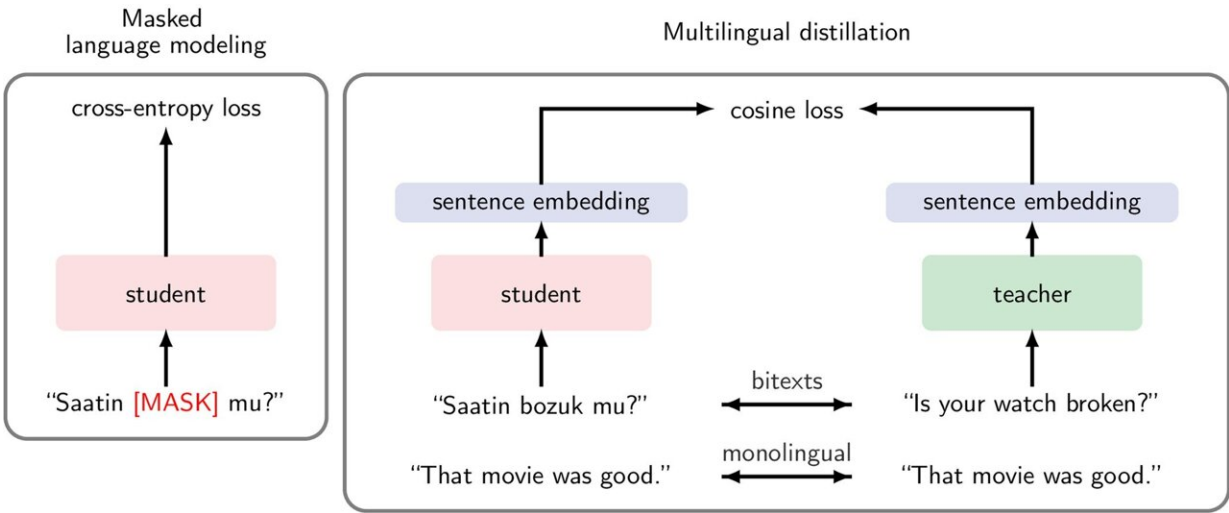# Meta's AI can translate dozens of under-resourced languages

June 7 2024

Architecture of the LASER3 teacher-student approach. Credit: *Nature* (2024). DOI: 10.1038/s41586-024-07335-x

The technology behind Meta's artificial intelligence model, which can translate 200 different languages, is described in a paper published in *Nature*. The model expands the number of languages that can be translated via machine translation.

Neural machine translation models utilize [artificial neural networks](#) to translate languages. These models typically need a large amount of accessible data online to train with, which may not be publicly, cheaply, or commonly available for some languages, termed "low-resource languages." Increasing a model's linguistic output in terms of the number of languages it translates could negatively affect the quality of the model's translations.

Marta Costa-jussà and the No Language Left Behind (NLLB) team have developed a cross-language approach, which allows neural [machine translation](#) models to learn how to translate low-resource languages using their pre-existing ability to translate high-resource languages.

As a result, the researchers have developed an online multilingual [translation](#) tool, called NLLB-200, that includes 200 languages, contains three times as many low-resource languages as high-resource languages, and performs 44% better than pre-existing systems.

Given that the researchers only had access to 1,000–2,000 samples of many low-resource languages, to increase the volume of training data for NLLB-200 they utilized a language identification system to identify more instances of those given dialects. The team also mined bilingual textual data from Internet archives, which helped improve the quality of translations NLLB-200 provided.

The authors note that this tool could help people speaking rarely translated languages to access the Internet and other technologies. Additionally, they highlight education as a particularly significant application, as the model could help those speaking low-resource languages access more books and [research articles](#). However, Costa-jussà and co-authors acknowledge that mistranslations may still occur.

More information: Scaling neural machine translation to 200

languages, *Nature* (2024). DOI: 10.1038/s41586-024-07335-x