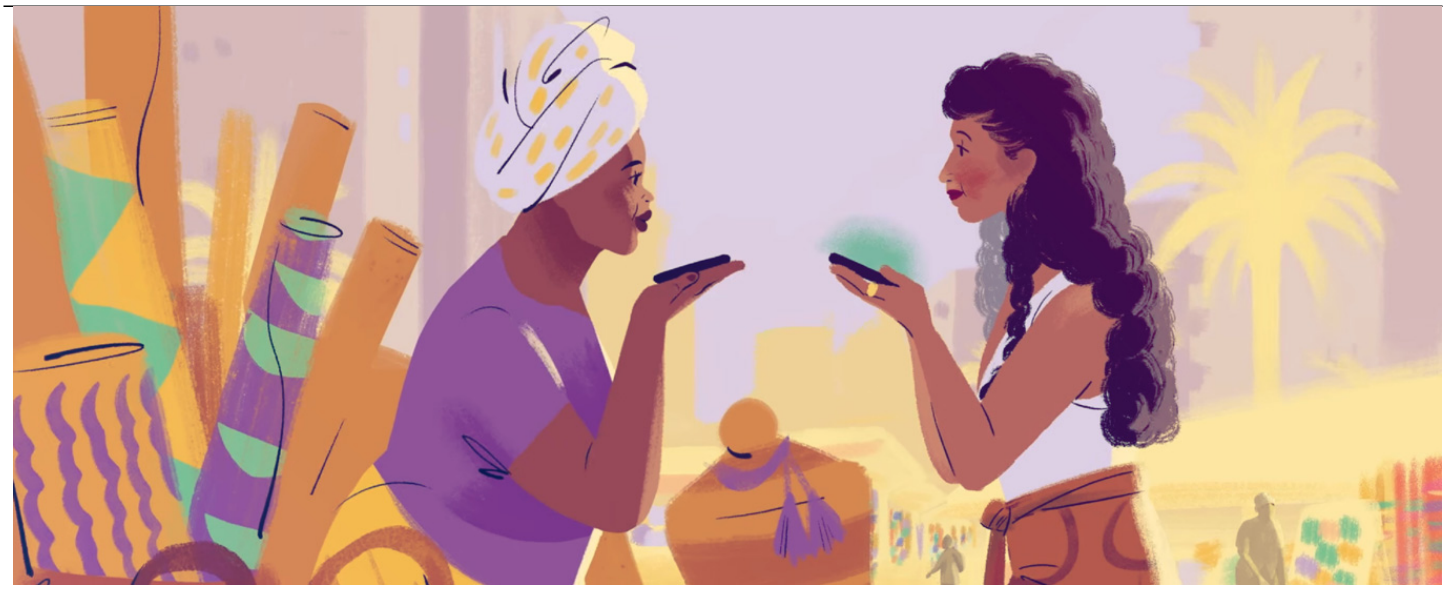
Meta ??????????????

Meta ??????????????

Centific ??



Artificial intelligence (AI) must be inclusive to reach its potential. AI applications that solve problems for a small segment of the population will fail to achieve widespread adoption. So, it's important that AI applications be designed and prepared with data that reflects as many segments of the global population as possible. Many moving parts need to be managed well to do that, and one of them is language. The more languages an AI application can handle, the more inclusive it is.

To that end, Meta took a step in the right direction by announcing recently that the company has created a single AI model capable of translating across 200 different languages -- including many not supported by current commercial tools. Meta is open sourcing the project in the hopes that others will build on its work.

What Meta Announced

For context, in February 2022, Meta announced that its AI researchers had created No Language Left Behind (NLLB), an effort to develop high-quality machine translation capabilities for most of the world's languages. What Meta announced in July was a breakthrough in NLLB: the building of a single AI model called NLLB-200, which translates 200 different languages with results far more accurate than what previous technology could accomplish.

NLLB-200 makes current technologies accessible in a wider range of languages, and in the future will help make virtual experiences more accessible, as well. NLLB improves the quality of translations across Meta's technologies by an average of 44%. For some African and Indian-based languages, NLLB-200's translations were more than 70% more accurate.

Meta also said it built FLORES-200, a dataset that enables researchers to assess NLLB-200's performance in 40,000 different language directions. FLORES-200 allows Meta to measure NLLB-200's performance in each language to confirm that the translations are high quality.

To help other researchers improve their translation tools and build on Meta's work, the company is opening NLLB-200 models and the FLORES-200 dataset to developers, in addition to its model training code and code for re-creating the training dataset.

Interestingly, Meta tied the effort to the evolving metaverse. In a statement, the company said, "As the metaverse begins to take shape, the ability to build technologies that work well in a wider range of languages will help to democratize access to immersive experiences in virtual worlds."

This important theme has emerged throughout Meta's public-facing content, for example here. The emphasis on inclusion is consistent with Centific's own belief that the metaverse must be mindful in order to succeed.

Our Take: AI Must Be Mindful

Making AI more inclusive spans the entire world beyond the metaverse. And it's a major challenge among many. In fact, AI needs to overcome numerous challenges to be mindful, such as:

- Bias being built into AI-based products.
- A lack of inclusiveness (the challenge Meta and many others are attempting to tackle). It should be noted that inclusivity in AI is more than a language issue.
- Consumer trust issues.

We define Mindful AI as follows: developing AI-based products that put the needs of people first. Mindful AI considers especially the emotional wants and needs of all people for which an AI product is designed – not just a privileged few. When businesses practice mindful AI, they develop AI-based products are more relevant and useful to all the people they serve.

To be mindful, AI must be human-centered, responsible, and trustworthy. Meta's announcement shows just how much work needs to be done in the area of inclusivity alone. Meta is casting its net far and wide by including "low-resource languages" in its model — languages with fewer than 1 million publicly-available translated sentence-pairs. These include many African and Indian languages not usually supported by commercial machine translation tools.

Training data to translate these languages can be enormously difficult, especially for more obscure idioms. And yet it must be done. As we've seen with the development of life-saving products such as pandemic vaccines, global problems are not close to being solved unless they treat every corner of the world. When a segment of a population is overlooked, the entire world is vulnerable.

As The Verge noted in response to Meta's news, language translation can be fraught with bias, too:

> Translation is a difficult task at the best of times, and machine translation can be notoriously flaky. When applied at scale on Meta's platforms, even a small number of errors can produce disastrous results — as, for example, when Facebook mistranslated a post by a Palestinian man from "good morning" to "hurt them," leading to his arrest by Israeli police.

So, how to ensure that languages are both inclusive and as free from bias as possible? The answer comes down to training data with checks and balances in place. For instance, at Centific, we keep people in the loop to train AI-powered applications with data that is unbiased and inclusive. We rely on globally crowdsourced resources who possess in-market subject matter expertise, mastery of 200+ languages, and insight into local forms of expressions such as emoji on different social apps. They help us prevent bias from creeping into the data sets we train.

Contact Centific

Mindful AI is not a solution. It’s an approach. There is no magic bullet or wand that will make AI more responsible and trustworthy. AI will always be evolving by its very nature. But Mindful AI takes the guesswork out of the process. Contact Centific to get started.

Image source: Meta