

AI for everyone in every language

15.9.2026 - James Manyika | Google Italia

We're moving beyond traditional text translation to build models that understand the world's rich, living languages exactly as they are expressed.

Today, our technologies and products power everyday interactions in more than 300 languages, spoken by more than 7 billion people — representing 86% of the global population. Reaching this milestone is meaningful, but it also underscores work that is critical to our mission. For decades, technology has worked best for a handful of dominant languages, leaving thousands of living languages and dialects poorly represented or absent altogether from the digital world.

When we launched Google Translate in 2006, our goal was simple: to break down the barriers between languages. Advances in AI have helped us bring that vision to more people, expanding Translate from a handful of languages to more than 250 today. But translating text isn't enough. Technology needs to understand how people actually communicate in the real world. So we focus our research and development on building systems that honor cultural nuance and the richness of human language, enabling everyone to participate and be understood on their own terms.

Here's what that work looks like in practice.

Going from text to true understanding

Historically, speech recognition systems followed a rigid, multi-step process: transcribing audio into text, processing that text, and then synthesizing it back into audio. While functional, this pipeline strips away the richest parts of human communication: tone, pacing, emotion, and context.

People don't speak in perfectly neat, grammatical sentences. We laugh, overlap, hesitate, and weave multiple languages together mid-sentence, like when we speak Spanglish or Hinglish.

To capture this, we moved beyond text transcripts to native audio intelligence — training models like Gemini to process audio directly as is, while also grasping both sound and intent. These efforts include:

- Fluid real-time dialogue tools:
 - Today, Gemini 3.5 Live Translate powers real-time spoken translation across 70 languages and 2,000+ language pairs, naturally capturing code-switching and emotional cues along the way.
 - Gemini 3.5 Transcribe is our most precise speech-to-text model yet, turning raw audio into polished, formatted text, even in noisy environments or with complex jargon. It also powers features like Rambler on Android Gboard, which removes filler words, fixes grammar and punctuation, and lets you edit or rewrite with voice commands and switch seamlessly between languages.
- The 1,000 Languages Initiative: AI is helping us break down language barriers at a scale that was previously unimaginable. But reaching more people in their preferred language means going beyond the languages where AI performs best today: Our goal is to support the world's 1,000 most-spoken languages. To help make that possible, our Universal Speech Model — trained on 12 million hours of audio — used cross-lingual transfer learning, techniques that enable models to transfer what they learn from data-rich languages, to improve speech understanding in languages with far less training data. This allows models to apply patterns

learned from data-rich languages to under-resourced ones.

- **Rigorous foundational research:** This work builds on 25 years of open research and more than 400 peer-reviewed speech papers, which have helped push the frontier and advance speech models.

Putting communities at the heart of language data

Because the web disproportionately represents a few dominant languages, teaching AI to understand underrepresented languages required us to rethink how we gather data. The solution is local grassroots partnerships. This localized approach has driven three of our most ambitious open-data partnerships:

- **WAXAL** (Wolof for “speaking,” pronounced “Wah-hal”): Built with partners including Makerere University and Digital Umuganda, WAXAL is a large-scale, open speech dataset covering 27 Sub-Saharan African languages spoken by more than 100 million people across more than 26 countries, capturing tonal variation and conversational rhythms often missing from traditional datasets.
- **Project Vaani:** In partnership with the Indian Institute of Science (IISc) and Bhashini, Project Vaani is mapping India’s linguistic diversity through a region-anchored rather than language-anchored approach, enabling it to collect to date more than 30,000 hours of speech across 109 languages from more than 155,000 speakers.
- **Amplify Initiative:** We teamed up with more than 1,600 local experts and 20 universities across four continents, including Brazil’s UFMG, India’s IIT Kharagpur, and Uganda’s Makerere University, to contribute 15,000 multimodal data points capturing local nuance.

We’re also building on our work prioritizing open-source language innovation through our new tool Language Explorer. It’s an interactive tool that visualizes LinguaMeta, the world’s largest open-source language data repository. Recognized by Fast Company for design innovation, it continuously maps more than 7,000 spoken, written, and signed languages.

The impact of these innovations and partnerships is greatest when they reach the people who can turn new data and insights into meaningful change in their communities. Google.org-supported efforts, including the Centre for Digital Language Inclusion and AI Singapore’s Project Aquarium, are helping bring multilingual tools to farmers, healthcare workers, teachers, and other essential community members around the world.

Overcoming real-world constraints

For more than 3 billion people ¹, reliable internet access is still out of reach. Technology is only truly accessible if it works where people live, including areas with limited or intermittent connectivity.

To help address this, we developed TranslateGemma, a family of lightweight open translation models built from Gemini and trained across 55 languages. Because TranslateGemma runs efficiently on-device, high quality translation no longer requires a connection to the cloud or the internet.

Still, running powerful AI models requires capable hardware, which excludes the hundreds of millions of people still using feature phones in low-resource regions. To bridge this divide, we’re supporting organizations like Viamo to power “Ask Viamo Anything” (AVA), a voice AI assistant that brings the power of Gemini to standard feature phones. Viamo has successfully piloted AVA in Rwanda with its existing interactive voice response users, and the service has already used Gemini

to answer more than 2 million questions.

Designing for accessibility

Language isn't just about regional dialects or vocabulary. It's also about the many other ways people communicate. Conventional speech tools frequently fail people with non-standard speech, making them adapt to the technology rather than the other way around.

We're working to change that by designing for accessibility from the ground up, for example, with Sign Language-to-Text (SL2T). Trained across 50+ sign languages, SL2T powers sign-to-text dictation in Gboard and Live Transcribe on Pixel 11, starting with American Sign Language (ASL) to English. This is an important first step toward making our products more accessible to the 70 million people worldwide who rely on sign language to communicate.

Getting local pronunciation right

Details matter, certainly in everyday tools like navigation. When a navigation app mispronounces a town or street name, it doesn't just cause confusion, it can overlook the cultural heritage of the place.

We believe cultural context should be part of how language technology is built, and that means working directly with local communities. For example, in New Zealand, we worked with Māori language experts to improve the place name pronunciation in Google Maps, helping make the experience more accurate and genuinely local. Incorporating culturally authentic pronunciations directly into our text-to-speech models ensures technology reflects the language and heritage of the communities it serves.

Building language tools for everyone

One essential lesson we've learned over 20 years of AI language research and development is that technology should never narrow the spectrum of human expression — it should expand it.

Today, our language technologies are already embedded across our core ecosystem, connecting more than five billion people across nine platforms, including Search, Android, Chrome, YouTube, and Google Play. But scale is only part of the story. The bigger goal is depth and richness — building systems that grasp context, respect culture, and celebrate the many ways people communicate.

As we expand our use of AI to address some of society's biggest opportunities, we'll continue to work closely with local communities to build technology that helps more people communicate, participate, and be understood on their own terms.

<https://blog.google/innovation-and-ai/technology/ai/ai-for-every-language>