INTERVIEW

Why sovereign AI must be local to truly work

James Kaplan, Co-Founder and CEO of MeetKai, explains why one-size-fits-all foundation models fall short, how language and cultural context reshape AI performance, and why sovereign, country-specific AI stacks are essential for education, healthcare, and public services.

JAMES KAPLAN
CO-FOUNDER AND CEO OF MEETKAI

How does MeetKai's sovereign AI stack differ from commercially available cloud or foundation model offerings?

The opportunity lies in addressing each market's needs in very specific ways. If you look at conventional, foundational LLMs available in the cloud today, whether from Azure, OpenAI, or Google Cloud, they are largely one-size-fits-most solutions.

That limitation becomes even more pronounced when language and cultural context are considered. Take Pakistan, for example; there are four or five widely spoken languages within a single country. A common challenge we see is that many LLMs treat Arabic as a single, uniform language, when in reality it is not. Dialects, usage, and context vary significantly across regions.

This is where our approach differs. We take localization seriously. Rather than offering a generic "American" or "Chinese" model and hoping it works everywhere, we focus on developing country-specific and context-aware solutions that reflect local languages, cultures, and real-world requirements. That specialization is central to how we deliver value across different markets.

What are the hardest AI challenges when building models for under-resourced languages?

One of the biggest challenges lies in evaluation and benchmarking. When building an LLM, there are essentially two parts to the problem. First, can you train a model in a particular language—for example, Kazakh? Technically, that is achievable, and we've seen examples where LLMs have been trained for languages like Kazakh.

The real challenge comes next: how do you ensure the model actually performs well for that language and for the intended use cases? This is where things become difficult.

If you look at how companies like OpenAI or Google release new LLMs, they typically publish performance results across 20 or more benchmarks. However, those benchmarks are largely English-centric, designed around English-language tasks, patterns, and use cases. They do not translate well to local or low-resource languages.

When working with country-specific languages, it's not enough to build a model that simply understands the language. You also need the ability to test, assess, and evaluate that model in the local language.

That requires deep domain expertise within those countries—not just raw data collection.

Collecting raw data is often the easiest part. The more complex challenge is digitization and contextual relevance. In many of these markets, there aren't readily available digital datasets or PDFs of textbooks. In many cases, the source material exists only as physical textbooks, which must first be scanned, digitized, and processed before they can even be used for training or evaluation.

This digitization gap was one of the earliest challenges highlighted to us and remains a critical barrier when building effective, high-quality language models for resource-constrained environments.

MeetKai promises to deliver sovereign AI models and agents that work in any language and across a variety of hardware platforms. How do you achieve that without compromising performance or accuracy?
It really comes down to specialization. There's a nuanced challenge around accuracy, because accuracy depends on how you measure performance—and that brings us back to benchmarking and evaluation. You have to take a scientific approach to assessing how these models perform.

To give you a concrete example, we are currently developing an Urdu language model. One interesting challenge we've encountered is that when a user asks a math or logic question in Urdu, you don't necessarily want the model to reason in Urdu. In many cases, the reasoning should happen in English, even though the response is delivered in Urdu.

This reflects how people actually think and learn in countries like Pakistan. While users may speak Urdu conversationally, mathematical and technical reasoning is often framed in English. For example, certain mathematical concepts—like "square root"—don't have direct equivalents in Urdu. Understanding these cognitive and linguistic nuances is critical.

This is why accuracy is not as simple as applying existing benchmarks. In English, there are well-established benchmarks for tasks like mathematical reasoning. In Urdu, those benchmarks don't exist. So our focus becomes: how do we design the right benchmarks and evaluation frameworks for each language and context?

That's where specialization matters most.

Rather than relying on generic benchmarks, we invest in creating language- and country-specific evaluation methodologies. This is what allows us to maintain both performance and accuracy across diverse languages and environments.

Which sectors do you think will feel the impact earliest—education, healthcare, agriculture, public services, or enterprise?
Based on what we've seen in our previous work, education and healthcare stand out as the earliest and most impactful areas, healthcare in particular.

In many of these countries, healthcare is absolutely critical to expanding access to essential services. One of the most powerful things AI does is expand access to expertise at a much lower cost. Traditionally, even models like telehealth still rely on a doctor at the other end of the connection. While that reduces costs compared to in-person care, it can still be prohibitively expensive at scale.

With AI-driven healthcare solutions, the economics change fundamentally. Instead of reducing costs from, say, 100 to 70, you're often starting from close to zero. That shift enables entirely new access models, especially in underserved or resource-constrained environments. It's not just about efficiency, it's about access expansion, and that's where healthcare, in particular, will see some of the earliest and most meaningful impact.

How do you define sovereign AI, and why is it such a critical concept for governments

"This digitization gap was one of the earliest challenges highlighted to us and remains a critical barrier when building effective, high-quality language models for resource-constrained environments."

and regulated industries today?
At its core, sovereignty is about control—and that is a very loaded term. When governments look at critical systems and national priorities, control over how citizens interact with technology and with the world is non-negotiable.