iKala Advances TMMLU+ to Strengthen Taiwan's Sovereign AI Evaluation Capabilities

Traditional Chinese benchmark expands toward agentic and multimodal evaluation, helping enterprises from measuring what AI knows to validating what AI can do.



"Taiwan should not only have access to the world's most advanced models. We also need the ability to evaluate them using our own data and standards," said Cheng.

AL, TAIWAN, September 15, 2026 /EINPresswire.com/ -- As the global AI race shifts from building larger models to deploying AI in real-world applications, iKala, a leading AI transformation company, is advancing TMMLU+, its open-source Traditional Chinese AI benchmark, to help Taiwan strengthen its ability to independently evaluate and define AI capabilities.

Sega Cheng, Co-founder and Chairman of iKala, believes sovereign AI should extend beyond building domestic foundation models. Instead, it should encompass four critical layers: compute, models, data, and applications—and, increasingly, the ability to determine whether AI systems actually meet local needs.

> Taiwan should not only have access to the world's most advanced models. We also need the ability to evaluate them using our own data and standards."
>
> *Sega Cheng, Co-founder and Chairman of iKala*

Building the Yardstick, Not Another Model

In 2023, as Taiwan's public sector, industry, and academia accelerated investment in foundation models, iKala took a different approach.

"If everyone is building models, who defines whether those models are actually good?" Cheng said.

That question led to TMMLU+, a large-scale benchmark designed to evaluate the multitask language understanding capabilities of large language models (LLMs) in Traditional Chinese.

TMMLU+ v1 contains 22,690 multiple-choice questions across 66 subjects, spanning knowledge from elementary education to professional domains. Approximately six times larger than its

predecessor, TMMLU, the benchmark provides broader and more balanced coverage for evaluating Traditional Chinese language capabilities.
The research behind TMMLU+ was published at COLM 2024, and the benchmark has since been incorporated into evaluation efforts by international research teams, including Google and Meta.



TMMLU+ v1.1, released in September 2026



“We want TMMLU+ to become a shared foundation for defining AI capabilities and building trust across Taiwan’s AI ecosystem,” Cheng said.

“We made TMMLU+ open source because we wanted it to be used by the broader ecosystem,” said Cheng. “A healthy AI ecosystem needs shared standards that industry and academia can test, validate, and improve together.”
For iKala, developing such standards is also closely tied to data sovereignty. An AI system capable of speaking Chinese does not necessarily understand Taiwan. Local laws, public policies, cultural references, professional terminology, and everyday contexts require locally grounded data that cannot always be adequately represented by global datasets.

As AI becomes an increasingly important interface for search, decision-making, and information access, iKala believes building trustworthy Traditional Chinese datasets and evaluation frameworks will be essential to ensuring Taiwan maintains the ability to define what effective AI means in its own context.
From Evaluating Answers to Evaluating Actions
The need for independent evaluation is becoming even more important as AI moves from general-purpose chatbots toward specialized models and AI agents.

Gartner predicts that by 2030, up to 90% of generative AI solutions will use domain-specific models (DxMs). For enterprises, this means the challenge will increasingly shift from gaining access to AI models to determining which model or system is best suited to a specific business scenario.

Public benchmark scores alone may not provide that answer.
Enterprises can spend significant time and resources on proof-of-concept projects only to discover that a model performs differently in real-world workflows, lacks sufficient understanding of Traditional Chinese business contexts, or fails to meet local requirements.

To address this gap, iKala is extending its evaluation capabilities from LLM evaluation toward agent evaluation.
The next stage of evaluation will move beyond asking whether AI can generate the correct answer. It will assess whether an AI system can understand user intent, select and use tools correctly, process multimodal information—including text, images, and audio—and successfully complete end-to-end enterprise tasks.
This represents a fundamental shift in how AI performance is measured: from evaluating knowledge to validating execution.

For enterprises, the most suitable AI may therefore not be the model with the highest score on a global leaderboard. It may instead be the model or agent that delivers the strongest results against the organization's own data, workflows, requirements, and cost constraints.
TMMLU+ v1.1 to Further Improve Evaluation Reliability.

TMMLU+ v1.1, released in September 2026Huggingface / Github
iKala also released TMMLU+ v1.1 in September 2026, an updated version of the benchmark designed to improve evaluation accuracy, quality, and reliability.
The latest release includes a systematic review of all 66 subject areas, with particular attention to questions involving laws, regulations, public policies, and other time-sensitive information. Incomplete, corrupted, or otherwise invalid questions were corrected or removed, while questions with multiple plausible answers or ambiguous answer choices underwent additional review by domain experts. Questions for which a single defensible answer could not be established were excluded.

These updates are designed to reduce ambiguity and factual inconsistencies, ensuring that TMMLU+ remains a reliable and relevant benchmark as AI capabilities and the underlying knowledge environment continue to evolve.
"We want TMMLU+ to become a shared foundation for defining AI capabilities and building trust across Taiwan's AI ecosystem," Cheng said.
As the global AI landscape continues to change, iKala believes Taiwan's long-term advantage will not depend on any single foundation model. Instead, it will depend on maintaining control over the assets that allow the country and its enterprises to make informed AI decisions: local data, evaluation standards, and the ability to independently select, measure, and validate AI systems.

[About iKala]
iKala helps enterprises make better, faster decisions by embedding AI and data at the core of their business. We support AI transformation by helping organizations move from data to

decisions, delivering full AI solutions that combine their first-party data with iKala's intelligence built on billions of global social signals.
Headquartered in Taiwan with a global footprint, iKala serves over 1,000 enterprises and 50,000 brands across more than 190 countries, including Fortune 500 companies.
iKala Official Website: https://ikala.ai
Kolr Official Website: https://kolr.ai
Kuroma Official Website: https://kuroma.ai

Sunny Shih
iKala
+886 917 489 240
sunny.shih@ikala.ai
Visit us on social media:
LinkedIn

This press release can be viewed online at: https://www.einpresswire.com/article/941086168

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2026 Newsmatics Inc. All Right Reserved.