



MBZUAI's Institute of Foundation Models launches K2 Horizon, the world's largest fully open AI models in history

K2 Horizon Enables Greater Access and Reproducibility for Researchers and Developers Worldwide; Ready to Deploy from Edge to Enterprise

Abu Dhabi, UAE, 3 September 2026: The Institute of Foundation Models (IFM) at the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) today launched K2 Horizon, a fleet of six AI foundation models ranging from 0.9 billion to 375 billion parameters. Every model is released fully open, with weights, code, training data, and methodology, allowing researchers, developers, and institutions worldwide to inspect, reproduce, and adapt them.

Across reasoning, mathematics, coding and agentic tasks, K2 Horizon delivers top-tier performance in every size class, with the 0.9B, 3.7B and 7B models setting new state of the art at their respective scales. The 0.9B model is designed for highly constrained environments such as watches and glasses, while the 3.7B and 7B models bring advanced capabilities to phones and other on-device applications. The dense 32B model and sparse 36B-A4B model provide powerful options for local hosting and on-premise servers, and the 375B-A23B model brings the fleet's strongest capabilities to demanding enterprise deployments.

K2 Horizon is the largest fully open model release in AI history, and it cements the UAE's leadership in building frontier-grade AI systems that are developed by national institutions and shared openly with the world. The launch gives developers and researchers greater choice in capabilities, scale and deployment options. While most releases described as "open" stop at publishing weights, the K2 family has moved progressively toward end-to-end transparency, a trajectory that began with K2, continued through K2-Think and K2 Think V2, and reaches its widest expression with K2 Horizon.

The six models share a core architecture, vocabulary, training methodology, interfaces and deployment tooling, with a smaller vocabulary for the 0.9B model. IFM's dynamic model routing technique directs tasks to the most cost-effective model and provides developers with a practical path from prototype to production. The fully open code, training data and recipes are a significant step forward in transparency, well beyond the "open weights" dialogue that has dominated AI industry headlines over recent weeks.

Professor Eric Xing, Founder of IFM, and President and University Professor of MBZUAI, said: "We believe meaningful AI progress depends on the ability to examine, build upon, and improve the technology, not simply access it through an API. The most important technology of our time should be built with the world, not kept from it. K2 Horizon is what that conviction looks like when an institution



commits to it completely, and it reflects the environment we work in here in the UAE: a country that decided to build AI rather than wait for it.”

K2 Horizon includes key technology innovations: 1) A unique "diffusion distillation" technique that generates blocks of tokens in parallel, which speeds model performance by roughly 3X without degrading response quality and 2) A "mixture of value attention" architecture that improves reasoning without adding more computation; and 3) Dynamic model routing, which directs tasks to the most cost-effective model and provides developers with a practical path from prototype to production.

K2 Horizon's three smallest models achieve state-of-the-art results in their respective classes across mathematics, reasoning, software-engineering and agentic capabilities. Model highlights:**0.9B**: the industry's best-performing model of its size in mathematics, reasoning and tool use; compact enough to run on a watch, and suited to edge deployment in sectors including energy, logistics, healthcare and public services.

- **3.7B**: a fine-tuning-friendly size for developers and researchers, with reasoning that is the industry's best under 4B and matches or exceeds many larger models.
- **7B**: the industry's best-performing model under 10B parameters, with strong software engineering and deep research capability; small enough to run on a phone.
- **32B**: among the most capable dense models for local hosting; built for heavy use on laptops and on-premise servers.
- **36B (4B active)**: a new Mixture of Value Attention (MoVA) architecture that outperforms many larger models while activating only 4B parameters.
- **375B (23B active)**: the flagship, built to compete with leading open-weight models on reasoning and agentic work, and enterprise-ready.

Each K2 Horizon model is engineered to compete with the best open models of its size on reasoning, tool use, and agentic tasks. The smallest is built for devices where memory, power consumption and local processing matter; the largest for demanding reasoning and enterprise workloads. Developers can prototype on one and scale to another without changing model architecture or workflows.

Hector Liu, Director of IFM's Silicon Valley Research Lab, added: “Rather than release a single model, we're releasing an entire fleet at once, six models, from one small enough to run on a watch to a flagship built for enterprise reasoning. Every model is built to compete with the best open models at its size, and everyone ships with the weights, code, training data, and methodology behind it. Developers can prototype on the smallest model, scale to the flagship, and verify every claim we make along the way.”

The launch of K2 Horizon also marks IFM's continued growth as a global AI research lab operating from Abu Dhabi. Established in May 2025, IFM now employs AI researchers, scientists and developers across three offices in Abu Dhabi, Silicon Valley and Paris, a distribution that places frontier model development in the UAE while drawing talent from the world's leading AI centers.

Today, MBZUAI is an integral part of the UAE's AI transformation, training 80,000 federal employees as agentic AI experts under a strategic knowledge partnership with the Federal Authority for



Government Human Resources. K2 Horizon extends that role into research: models built for reasoning, tool use, and agentic tasks.

K2 Horizon models are available immediately through [Hugging Face](#), as well as vLLM and SGLang. The API of K2 Horizon is available today through IFM's global ecosystem of inference partners, including Compass, Cerebras, AWS and Nebius. K2 Horizon models and code are released under the Apache 2.0 license.

ENDS

About the Institute of Foundation Models (IFM)

The [Institute of Foundation Models](#) (ifm.ai) is a global AI research lab dedicated to open and independent development of frontier-class foundation models. Launched in May 2025 by MBZUAI, IFM operates from advanced lab facilities in Abu Dhabi, Silicon Valley and Paris. IFM's model portfolio includes K2 series, the fully open-source fleet of LLMs; Jais, the most advanced Arabic LLM; and PAN, the next-gen World Model for embodied reasoning and real-world simulation.

About Mohamed bin Zayed University of Artificial Intelligence

Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) is a research university in Abu Dhabi. Founded in 2019, it enrolled its first graduate students in 2021 and opened a highly selective undergraduate program in 2025. Recognized globally for its research excellence, it has nearly a thousand students from more than 70 countries. Learn more at [mbzuai.ac.ae](#).