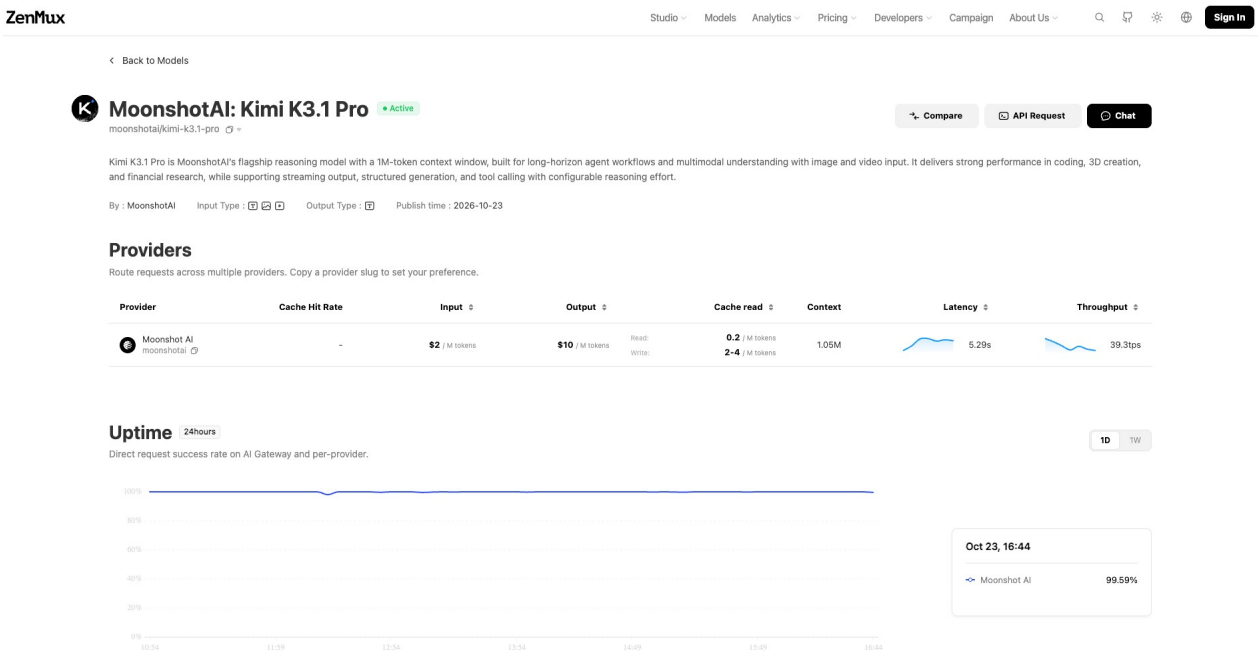


ZenMux Introduces Kimi K3.1 Pro for Long-Context AI Workflows

October 29, 2026

Singapore - October 29, 2026 -

ZenMux announces the addition of Kimi K3.1 Pro and Kimi K3.1 Pro Highspeed, expanding its model offering with two options from Moonshot AI's latest release. ZenMux Kimi K3.1 Pro is now listed as active alongside the Highspeed variant, giving development teams a choice of models for coding, spatial reasoning and multi-step agent workflows that require extended context and sustained work across connected tasks.

Moonshot AI's official announcement describes K3.1 Pro as the first model in its K3.1 series and a post-trained version of Kimi K3. The developer reports a one-million-token context window and improvements across software engineering, 3D creation and research tasks. These specifications describe the underlying model; application behavior also depends on the tools, instructions and execution environment connected to it.

The release places particular emphasis on work that continues beyond a single response. In software development, that includes examining an existing project, making changes, running tests and revising an

implementation after reviewing the results. Moonshot AI presents these longer sequences as a central part of the model's development, alongside shorter coding tasks such as fixes and individual feature requests.

For teams evaluating the model, the larger context window provides a basis for testing workflows involving substantial project material. Repositories, documentation and prior task history can form part of those evaluations, subject to the constraints of the application and API request. Context capacity alone does not establish that every supplied detail will be interpreted correctly, so task completion and output quality remain separate assessment criteria.

Moonshot AI also highlights 3D modeling and spatial reasoning in its release. Its examples include generating CAD code from visual references and writing scripts for tools such as Blender to construct or modify assets. The announcement describes work involving shape, proportions, geometry, materials and scene layout, extending the release's scope beyond conventional text and code completion.

Those demonstrations involve a model operating with software tools and execution support, rather than a claim that every API response directly produces a finished 3D asset. The distinction matters for developers integrating the model into an existing product: a usable workflow may require a renderer, file handling, script execution and checks against the original design brief. Tool access and validation remain part of the surrounding system.

The official release also addresses source-grounded research and work across multiple documents. Moonshot AI reports improvements in tasks that require retrieval, cross-document verification and numerical reasoning, with results measured through both external and internal evaluations. These findings provide context for the model's intended uses, but they are developer-reported results rather than independent measurements of performance through ZenMux.

ZenMux Kimi K3.1 Pro Highspeed is listed as the faster variant, with approximately two to three times higher output speed and comparable intelligence, according to the model description and Moonshot AI's announcement. Both ZenMux listings support text, image and video input with text output and a context length of 1,048,576 tokens. The published speed comparison describes the variant's positioning, not an independently measured latency guarantee for every request through ZenMux.

The model's release schedule also distinguishes hosted access from downloadable weights. Moonshot AI says the full model weights are scheduled for release by November 4, 2026. That timetable is separate from the company's announcement of access through its own products and API, and from the integration of the model into other platforms.

Within ZenMux's model offering, the two additions give development teams options to assess against their

own engineering and document-processing requirements. The standard Pro listing describes support for streaming output, structured generation and tool calling with configurable reasoning effort, while Highspeed is positioned for more responsive long-horizon workflows. Relevant evaluations can examine task completion, connected-tool behavior and output quality under each application's operating conditions.

Both models are listed in ZenMux's public model catalog under separate identifiers: moonshotai/kimi-k3.1-pro and moonshotai/kimi-k3.1-pro-highspeed. Their model pages provide distinct integration destinations, while Moonshot AI's technical announcement supplies the underlying capability descriptions and evaluation context. This separates the choice of model variant from the application tools and execution environment required to turn its responses into completed work.

For more information, visit https://zenmux.ai/moonshotai/kimi-k3.1-pro

About ZenMux:

ZenMux is an enterprise-grade large model aggregation platform with an insurance payout mechanism. The platform provides one-stop access to the latest models across providers. When issues such as poor output quality or excessive latency occur during use, our intelligent insurance detection and payout mechanism automatically compensates, addressing enterprise concerns around AI hallucinations and unstable quality
###
For more information about ZenMux, contact the company here:ZenMuxEmberember@zenmux.aiSingapore

ZenMux

ZenMux is an enterprise-grade AI model aggregation platform with a built-in insurance payout mechanism, offering one-stop access across providers and automatic compensation for quality or latency issues.

Website: https://zenmux.ai/

Email: ember@zenmux.ai

AI Simplified & Assured.

ZenMux