

## ZHIPU AI LAUNCHES GLM-5.1 CHINA'S FRONTIER MOE MODEL ON DOMESTIC CHIPS

Zhipu AI has launched GLM-5.1, a 744 billion parameter Mixture-of-Experts model that activates 40 billion parameters per token for efficient inference, and was trained entirely on Chinese compute infrastructure using Huawei Ascend 910B chips. On coding benchmarks, GLM-5.1 achieved 94.6% of Claude Opus 4.6's score, demonstrating frontier-level performance without access to restricted Western hardware signaling a strategic decoupling of the global AI supply chain.

GLM-5.1's full-stack domestic production from compute to model represents a milestone for Chinese AI sovereignty. The model's MoE architecture delivers high-speed inference without compromising reasoning depth, making it viable for enterprise agentic workloads. For SCBX and regional financial institutions, the emergence of multiple sovereign frontier AI models means greater optionality in model selection, reduced dependence on any single provider, and new considerations around data residency, regulatory compliance, and model governance as the competitive AI landscape rapidly globalizes.



WANT TO STAY AHEAD ON AI TECH FRONTIER,  
CONNECT WITH US  
E-MAIL: [RD@SCBX.COM](mailto:RD@SCBX.COM)

