

Personalized Learning. Scalable Education. Support Gen AI Tutoring at the Edge

Edge AI Solutions for Use in the Education Industry

Educational content is moving to digital formats, as students increasingly rely on online resources. Tools like SocratiQ, an AI-powered educational assistant, provide tailored explanations, adaptive assessments, and interactive conversations, enabling personalized learning at scale. However, SocratiQ leverages Large Language Models (LLMs) via cloud inference only. Cloud dependency has several disadvantages such as latency, bandwidth limitations, security risks, and high long-term costs.

Students in remote or low-connectivity environments were especially disadvantaged by this cloud dependency. Connectivity issues led to data leak, disengagement of students, and left teachers without timely insights. For personalized learning to be truly scalable, the solution needed to run at the edge. Utilizing Qualcomm Dragonwing™ AI On-Prem Appliance Solution powered by Qualcomm® Cloud AI 100 ultra accelerator enabled near-edge inferencing of SocratiQ, bringing real-time, security-focused, and cost-efficient AI tutoring directly into the classroom as an AI tutor in the box.

Dragonwing AI On-Prem Appliance Supports SocratiQ

Using the Dragonwing AI On-Prem Appliance with the Qualcomm Cloud AI 100 ultra accelerator in Aetina's MegaEdge AIP-FR68 AI workstation, SocratiQ can now deliver high-performance generative AI, natural language processing, and computer vision workloads directly at the edge.

Local hosting eliminates latency, lowers long-term costs, and provides quicker on-device access. Educators gain instant insights with sub-millisecond response times to adapt lessons in real time and keep students engaged. The solution scales to more than 40 users, supports models with up to 120B parameters, and enables contextual learning through RAG workflows while remaining compatible with OpenAI APIs for flexible, future-ready deployment.



Benefits

- **Offline access:** Learning continues without internet connection, essential for remote or bandwidth-limited regions
- **Enhanced security:** Student data stays local with on-prem inference, reducing exposure associated with use of cloud
- **Low latency:** Real-time responses ensure smooth, interactive tutoring experiences
- **High reliability:** Consistent performance with no disruptions from connectivity issues
- **Lower total cost of ownership:** Can help lower total cost of ownership compared to fully cloud-based deployments
- **Scalable performance:** Supports up to 120B parameter models and 40+ concurrent users for institution-wide learning
- **Educator insights:** Real-time analytics give teachers actionable visibility to track progress and adapt instruction

Use Cases

- **Scalable curriculum integration:** Can support generative AI across open-source and multilingual digital textbooks, enabling accessible learning for students worldwide
- **Personalized classrooms:** Adaptive quizzes, explanations, and interactive chat create tailored learning journeys
- **Student engagement:** Gamification with streaks, badges, and heatmaps keeps learners motivated and active
- **Educator support:** Knowledge graphs and analytics give teachers clear visibility into progress and comprehension

Powered by Qualcomm® technology

At the core of SocratiQ's near-edge AI solution is the Dragonwing AI On-Prem Appliance with Qualcomm Cloud AI 100 ultra accelerator, delivering high-performance inference for generative AI, NLP, and computer vision workloads.

- **Qualcomm Cloud AI 100 Ultra:** Enables local hosting of LLMs such as Llama3, Llama4, and Gpt-oss, eliminating cloud dependency
- **Hardware partner - Aetina:** MegaEdge AIP-FR68 AI workstation hosts the Qualcomm Cloud AI 100 Ultra supporting scalable, on-premise learning deployments

Ready to Redefine the Learning Experience?

Bring near-edge generative AI with personalized tutoring at scale into the classroom, powered by Qualcomm Technologies.

Contact sales to get started

