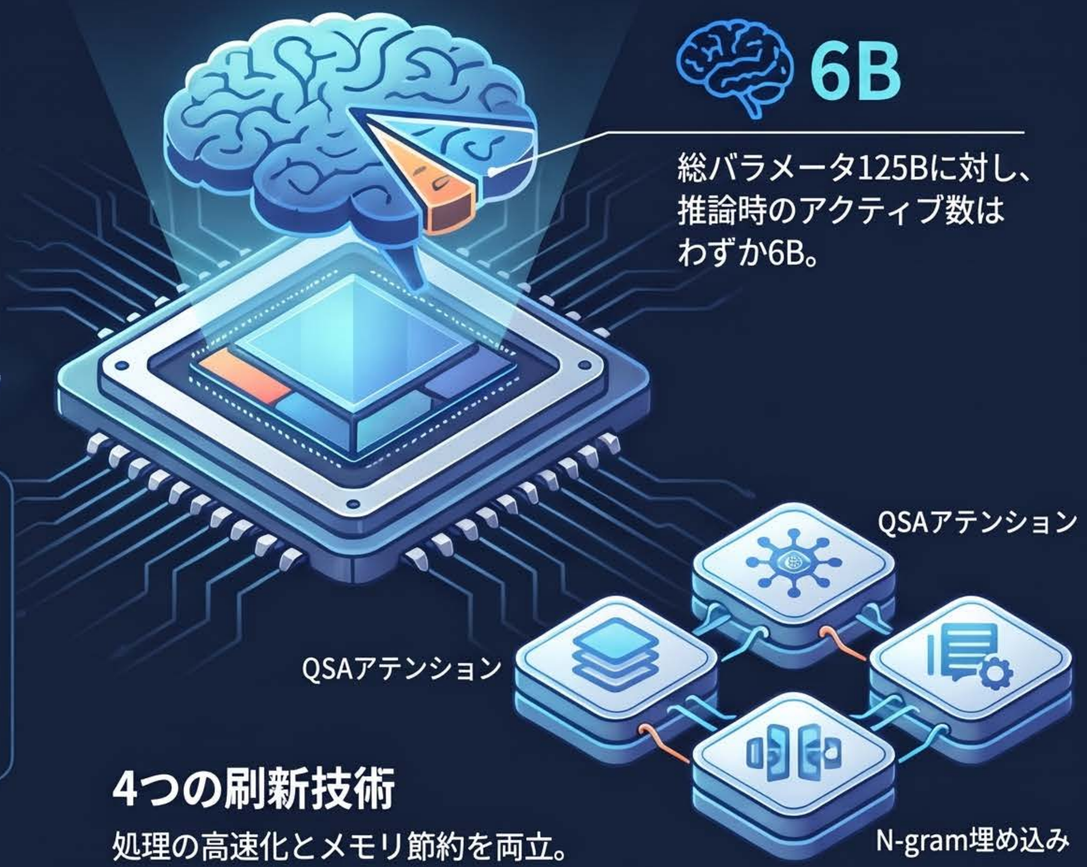
Qwen3.8-Flash-Next：次世代AI「Qwen4」の技術を先取りする高効率モデル

Qwen4アーキテクチャを先行採用した実験的プレビュー版。125Bパラメータながら6Bアクティブ数で、既存フラグシップ級の性能を低コストで実現する革新的なモデル。

Qwen4への系譜

Qwen4アーキテクチャの早期プレビュー：Qwen4の本格稼働前に、新構造をコミュニティで検証するための実験的モデル。

125B MoE（Mixture of Experts）構成



4つの刷新技術

処理の高速化とメモリ節約を両立。

圧倒的な効率性と実務への示唆

学習コストを約90%削減

前世代（Qwen3.7-Plus）比で、約9分の1の学習コストを達成。

Claude Opus 4.6を上回るベンチマーク

自社評価において、コーディングやオフィス業務の多くの項目で優位を主張。

主要ベンチマークにおける競合モデルとの性能比較（Alibaba公表値）

独自ライセンスへの留意

Apache 2.0ではなく「Qwen Community License 1.0」を適用。