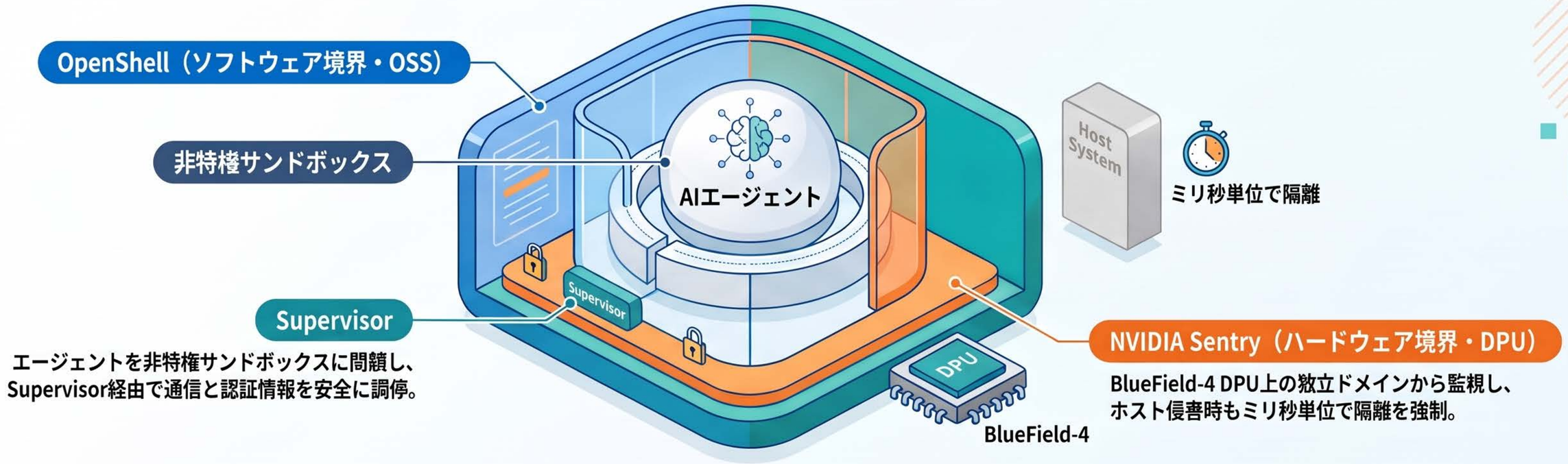
NVIDIA Open Agent Safety Platform：AIエージェントを外側から安全に封じ込めるアーキテクチャ



Supervisor

エージェントを非特権サンドボックスに閉鎖し、Supervisor経由で通信と認証情報を安全に調停。

NVIDIA Sentry（ハードウェア境界・DPU）

BlueField-4 DPU上の独立ドメインから監視し、ホスト侵害時もミリ秒単位で隔離を強制。

実装・運用における重要な注意点

デフォルト動作は「監査（audit）」モード



形式検証（Policy Prover）のカバー範囲



Gemini Notebook