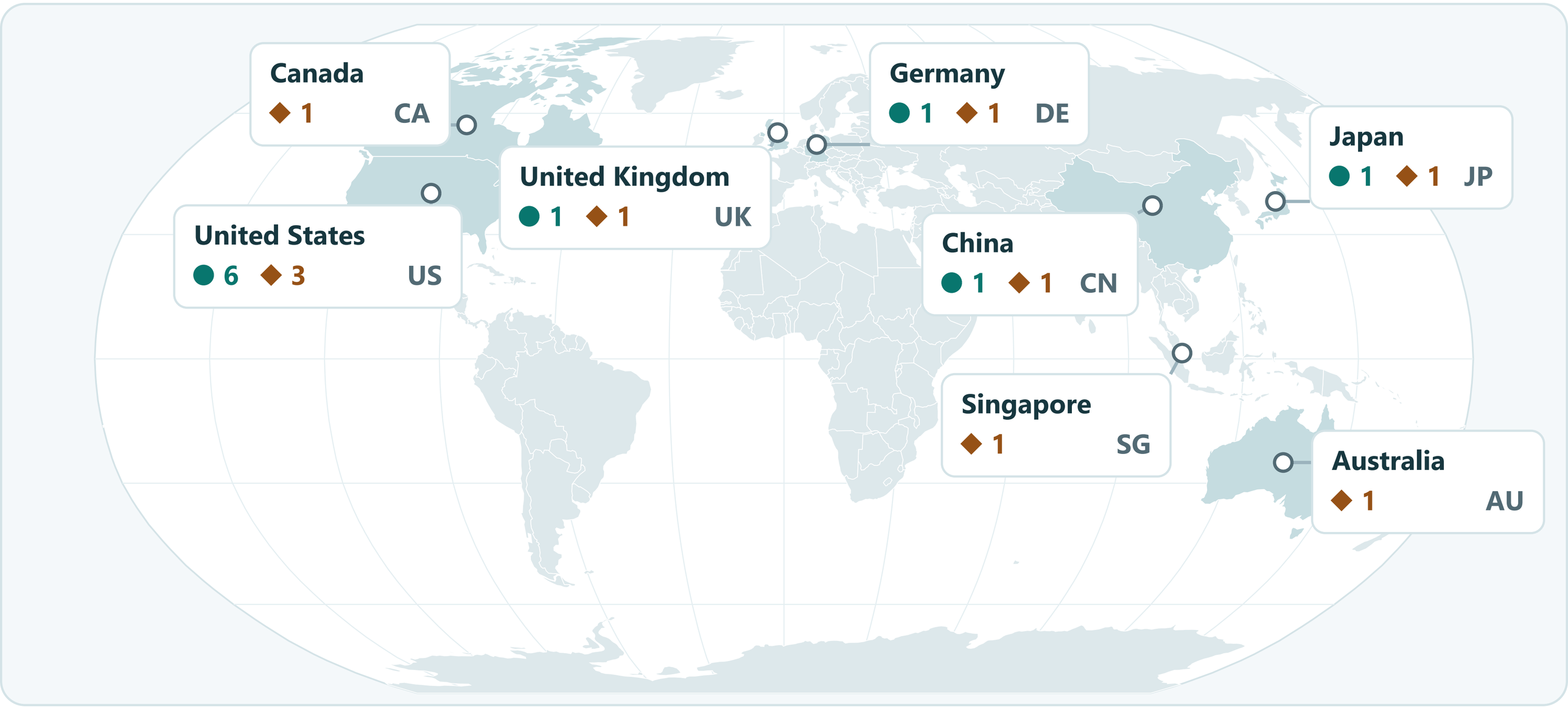
Geographic distribution of selected cases

20 organizations / 8 countries



● Commercial companies 10

CN	**Alibaba Group / Alibaba Cloud** Qwen3Guard live content screening	B A
US	**Amazon / AWS** AgentCore tool policy and guardrails	D B C
US	**Anthropic** Constitutional AI and frontier risk evaluation	A
UK	**Google DeepMind / Alphabet** Frontier safety thresholds and AP2 mandates	A D
US	**Meta** Scaling framework and Llama Guard	A B
US	**Microsoft** Foundry red teaming and Entra access	A D
JP	**NTT Group / NTT DATA** AI Assurance and AI Protection	A B
US	**NVIDIA** OpenShell policy and NeMo guardrails	D B C
US	**OpenAI** Preparedness and release governance	A
DE	**Spherity** Register-backed authority and runtime assurance	D C

◆ Research organizations 10

SG	**AI Verify Foundation** Moonshot application testing
AU	**Australian AI Safety Institute** Scalable oversight research
CN	**Beijing Academy of Artificial Intelligence** FlagEval model testing
DE	**Fraunhofer IAIS** Industrial AI assurance and assessment
JP	**Japan AI Safety Institute** AI agent evaluation guidance
CA	**Mila / Quebec AI Institute** Open safeguards and youth safety
US	**MLCommons** AILuminate safety benchmarks
US	**NIST CAISSI** Frontier evaluation and monitoring
US	**Stanford HAI / CRFM** AI Index and HELM Safety
UK	**UK AI Security Institute** Frontier testing and Inspect

A–D company activity key — A Primary — A Supporting — Country tags link rows to map counts.

A Improve and qualify behavior
B Observe and manage live behavior
C Design bounded execution
D Verify and enforce each action