40° South Guard

Guard alongside Azure AI Content Safety

Different problems, different products

Version 1.0 | May 2026

Why this comes up

Azure AI Content Safety is Microsoft's content moderation service, detecting harmful content across categories like violence, self-harm, sexual content, and hate speech. It also includes prompt shields for jailbreak detection.

The question we hear is: "We use Azure AI Content Safety to filter harmful outputs. Do we still need Guard?"

The short answer: content safety and regulatory compliance are separate disciplines. Azure AI Content Safety answers "does this output contain harmful content?" Guard answers "can we demonstrate to a regulator that our AI governance controls meet CPS 234, the Privacy Act, and the AI Safety Standard?"

Side by side

Capability	40 South Guard	Azure AI Content Safety
Primary purpose	Per-call compliance evidence and attestation for all AI providers	Harmful content detection (violence, self-harm, sexual, hate) and prompt shields
Architecture	Inline API proxy with compliance engine, attestation, and evidence vault	API-based content classifier. Called separately or integrated via Azure OpenAI.
Australian PII	Purpose-built detection with checksum validation for TFN, Medicare, ABN, and ACN	No Australian-specific PII types. US-focused PII detection only.
Regulatory mapping	Every call mapped to CPS 234, the AI Safety Standard, and the Privacy Act	No regulatory mapping. Content severity scores only.
Per-call attestation	Cryptographically signed attestation per API call with tamper-evident 7-year audit trail	Content classification results per call. No signed attestation or evidence vault.

AI-specific controls	Prompt injection detection, model inventory, shadow AI blocking, policy enforcement	Prompt shields (jailbreak detection), content classification across four harm categories
Data residency	Australian-hosted infrastructure. Data never leaves AU jurisdiction.	Depends on Azure region configuration. Customer responsibility to select AU East/Southeast.

Could you run them together?

Yes. Azure AI Content Safety is a good tool for keeping harmful content out of your AI outputs. Guard is the compliance layer that proves your governance controls are active.

A practical deployment would use Azure AI Content Safety to block harmful content in your Azure-hosted models, and Guard to generate regulatory evidence and attestation across all providers, including Azure, AWS Bedrock, Google Vertex, Anthropic, and any other API you call.

Azure keeps the content safe. Guard keeps the auditors satisfied.

Suggested next step

We can run a 30-minute proof of value showing Guard generating compliance evidence for AI calls that Azure AI Content Safety is already moderating, plus calls to non-Azure providers where you currently have no visibility.

Contact

hello@40south.au

40south.au

This document is confidential and intended for the named recipient only. May 2026.

Sources

1. APRA Prudential Standard CPS 234 Information Security, July 2019. www.apra.gov.au/sites/default/files/cps_234_july_2019_for_public_release.pdf
2. Microsoft Azure, "Azure AI Content Safety documentation." learn.microsoft.com/en-us/azure/ai-services/content-safety/
3. Australian Government, "Voluntary AI Safety Standard," August 2024. www.industry.gov.au/publications/voluntary-ai-safety-standard
4. OAIC, "Chapter 8: APP 8 — Cross-border disclosure of personal information." www.oaic.gov.au/privacy/australian-privacy-principles/australian-privacy-principles-guidelines/chapter-8-app-8-cross-border-disclosure-of-personal-information

5. 40 South Guard technical documentation. 40south.au