Trusted AI Boundary Canvas™

A one-page tool for defining what AI is allowed to do.

aska.net.au

1 Purpose, Outcome & Use Case

What work are we improving, what outcome matters, and what is AI here to help achieve?

> Define the workflow, user need and intended benefit.

2 Consequence Level

What happens if AI gets this wrong?

LOW | MODERATE | HIGH | CRITICAL

Lower consequence — *Higher consequence*

3 Access Boundary

What data, context, tools and systems may AI access?

> Use minimum necessary access.

What level of AI authority is justified by the evidence we have today?

Use this canvas to define boundaries, decision rights and the next experiment.

4 Authority Boundary

What may AI inform, assist, recommend or do?

- INFORM
- ASSIST
- RECOMMEND
- ACT WITH APPROVAL
- ACT WITHIN BOUNDARIES

Greater authority should be based on demonstrated competence — not technical capability alone.

5 Human Decision Rights

Who reviews, approves, overrides, intervenes and remains accountable?

> Define where human authority remains.

6 Evidence & Tolerances

What performance must be demonstrated? What failure is acceptable?

> Define measures, baseline, thresholds and demonstrated tolerance boundaries.

Predictable? • Detectable? • Containable?

7 Escalation & Containment

When must AI stop? How is failure detected, contained, reversed, escalated or handed back?

> Design safe fallback and response rules.

8 Next Experiment / Authority Decision

What will we test next — and what does the evidence justify?

EXPAND | MAINTAIN | RESTRICT | STOP

Start small. Learn deliberately.
Expand authority only when the evidence justifies it.

Governance makes the safe path the fast path.