Manifest

AI Ethics Handbook

Fair AI: the due diligence Manifest runs on every AI use case, before we build and for as long as it runs

THREE GATES

1 **Intake**

Describe the system, data, users and context

2 **Red-line check**

Stop anything Manifest will not build

3 **High-risk screen**

Six triggers decide the review path

Manifest · ISO/IEC 42001 certified · October 2026

Our commitment

Manifest

THE PRINCIPLE

Every AI use case passes the Fair AI Assessment before we build and stays under review for as long as it runs.

AI earns trust only when the people it affects can rely on it, and when our clients can show their regulators, boards and communities that it was built with care.

We do this due diligence for our clients, not after them. Each engagement produces evidence a risk owner can file: a FAIRA-ready risk record, an AI register entry, test results and a sign-off record.

Everyone who works on a Manifest AI system, whether employee, contractor or partner, applies this handbook in their daily work and raises a concern whenever something does not feel right.

This handbook is a practical guide to Manifest's AI ethics commitments. It does not replace legal, privacy or security advice, or a client's own governance processes.

About this handbook

Manifest

Who it covers

- Manifest employees and contractors
- Forward-deployed engineers
- Zeon AI Labs engineers on Manifest products
- Partners who resell or deliver Manifest solutions

What it covers

- Manifest aOS and every workspace on it
- Pre-built and custom agents in the Agent Hub
- Proofs of value, pilots and trials
- Third-party models Manifest embeds
- AI tools Manifest uses internally

How to use it

- Apply the ten principles to every use case
- Run each new use case through the Fair AI Assessment
- Work through the checklist for each delivery stage
- Hand over the evidence pack and keep it current

Its outputs drop straight into a client's own risk process, in particular Queensland's FAIRA framework and the national Guidance for AI Adoption.

Ten Fair AI principles

Manifest

Principles 1 to 8 follow Australia's AI Ethics Principles one for one. Principles 9 and 10 are Manifest's own.

1

Wellbeing

AI only where it brings clear benefit, with the smallest model that does the job

2

Human-centred values

Start with the people who use it and the people it affects

3

Fairness

Define fair for each use case; test before go-live, monitor after

4

Privacy and security

Privacy Principles, SOC 2 and ISO 27001; no training on client data without consent

5

Reliability and safety

Grounded in client documents, tested on real scenarios, red-teamed

6

Transparency

Users know when it is AI, and answers show their sources

7

Contestability

Anyone affected can ask a person to review the outcome

8

Accountability

A named person owns every use case; AI never does

9

Sovereignty

Data, models and audit trails stay in the agreed jurisdiction

10

Respect for communities

Consult the communities served and respect Indigenous data sovereignty

Governance and roles

Manifest

ESCALATION PATH

Manifest Board

Approves the handbook, decides escalations, gets a quarterly summary

AI Ethics Review Panel

CEO, CTO and AI Governance Lead; decides every high-risk case

AI Governance Lead (Naived Merchant)

Runs the process and the AI register; first contact for concerns

Use Case Owner

Owns one use case end to end, and its evidence pack

DELIVERY AND PARTNER ROLES

Engineers

Manifest and Zeon AI Labs engineers build and test to the agreed controls, and record data, models and results.

Designer

Designs the human checkpoints, AI disclosures and feedback paths users see.

Client risk owner

Co-signs the values assessment and sign-off and owns the system in the client's risk register.

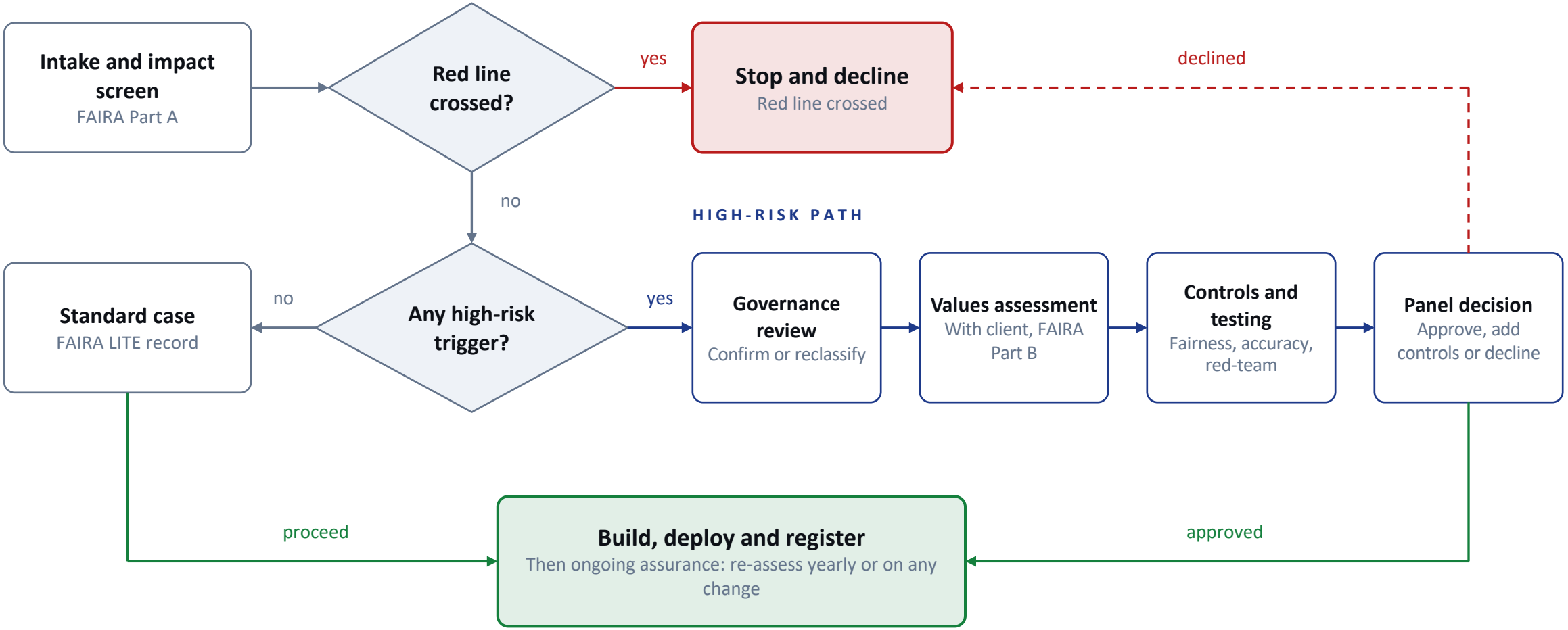
Partners

Apply this handbook to any Manifest solution they sell or deliver, and pass concerns on.

Everyone on an engagement completes Manifest's responsible AI training before starting and refreshes it every year.

The Fair AI Assessment

Manifest AI



Red lines: what Manifest will not build

Manifest

1

Personal freedom

Covert or mass surveillance, biometric targeting, social scoring, manipulation

Example
Matching CCTV footage to identify protesters

2

Harm to people

Foreseeable physical, financial or psychological harm; undermining elections or debate; excluding a group

Example
An agent that cuts benefits with no human review

3

Data sovereignty

Moving data or models out of the agreed jurisdiction; re-identifying people; training on client data without consent

Example
Sending a government's records to an offshore model without approval

4

Unlawful use

Breaching the Privacy Act, anti-discrimination law, the client's own law, or EU AI Act Article 5

Example
Inferring a person's religion or sexuality from their records

Found a red line at any stage? Stop work and tell the AI Governance Lead the same day.

Six high-risk triggers

Manifest

One yes makes a use case high risk. High risk means more care, not a refusal; most public-sector work meets at least one trigger.

1 Personal information

Does it process information about an identifiable person, in training, prompts or outputs?

2 Consequences for people

Could its outputs affect a person's rights, services, money, safety, health or job?

3 Automated decisions

Does it make or drive a decision without a human check before it takes effect?

4 High-risk sector

Government services, critical infrastructure, health, finance, employment, law enforcement or migration?

5 Agentic action

Is it an agent that acts by itself in systems of record such as SAP, TechnologyOne or Oracle?

6 Offshore or third party

Does data or a model leave the client's environment or jurisdiction, or go to a third party?

Three commitments: data, models, sovereignty

Manifest



Client data stays the client's

- Never used to train models for anyone else
- Fine-tuning for the client needs written consent
- Third-party models process only; they do not store or train on it
- Every agent action is logged with lineage and an audit trail



Models are chosen on evidence

- aOS is LLM-agnostic
- Each model is chosen on accuracy, cost, latency, bias and toxicity tests
- A change of model triggers a re-assessment
- Smaller models handle simple tasks, cutting cost and energy



The client decides where it runs

- Manifest cloud, private cloud or fully on-premises, including air-gapped
- Data residency agreed in writing and tested before go-live
- First Nations data follows the community's own rules

Checklists for each delivery stage

Manifest AI

1 Problem
Week 1

- ☐ Map friction and everyone affected
- ☐ Confirm AI is the right tool
- ☐ Intake and impact screen
- ☐ Check the red lines
- ☐ Answer the six triggers
- ☐ Name owners at Manifest and client

2 Ideation
Weeks 2 to 3

- ☐ Define and measure fairness
- ☐ Design the human checkpoint
- ☐ Design AI disclosure and feedback
- ☐ Agree data, residency and deployment in writing
- ☐ High-risk: joint values assessment

3 Solution
Under 2 months

- ☐ Ground agents; record lineage
- ☐ Test accuracy on real scenarios
- ☐ Test for bias
- ☐ Red-team prompts and actions
- ☐ Confirm audit log and shut-down path
- ☐ Evidence pack and joint sign-off

4 Operate
Ongoing

- ☐ Add to client and Manifest AI registers
- ☐ Monitor accuracy, fairness, feedback
- ☐ Report incidents on time
- ☐ Re-assess on any change, and every 12 months

The evidence pack every client receives

Manifest

9

records, ready to file in the client's own risk process

Each one maps to the Queensland FAIRA framework or a practice in the national Guidance for AI Adoption (AI6), so the client never starts its own assessment from a blank page.

Record	Produced at	Maps to
Intake and impact screen	Problem	FAIRA Part A
Risk classification and red-line check	Problem	FAIRA guideline, Appendix A
Values assessment (high-risk only)	Ideation	FAIRA Part B; AI Ethics Principles
Data card: sources, residency, lineage	Ideation	Privacy Act; principle 9
Model card: model, version, test results	Solution	AI6: test and monitor
Fairness, accuracy and red-team results	Solution	AI6: measure and manage risks
Human oversight and escalation design	Solution	AI6: maintain human control
Sign-off record and review dates	Solution	AI6: decide who is accountable
AI register entry	Operate	AI6: share essential information

Raising concerns and incidents

Manifest AI

Anyone can raise a concern about a Manifest AI system at any time. No one is penalised for raising one in good faith.

Staff and partners

Tell your Use Case Owner or the AI Governance Lead directly or email ethical.ai@manifestai.com.au.

Clients and affected people

Contact Manifest at ethical.ai@manifestai.com.au, or through the client. Concerns can be raised anonymously.

Incidents

Harmful output, bias, data leak, unsafe agent action or loss of control: stop or roll back, tell the AI Governance Lead within 24 hours, and tell the client as the contract requires.

Handbook review. The Board reviews this handbook every year, and sooner when the law or national guidance changes. Next review: 01 July 2027.

Reference frameworks

Manifest

Framework	Issued by	How Manifest uses it
FAIRA framework	Queensland Government	Structure of the intake (Part A) and values assessment (Part B)
FAIRA LITE	Queensland Government	Record for standard-case proofs of concept and trials
Australia's AI Ethics Principles	Australian Government	Principles 1 to 8
Guidance for AI Adoption (AI6)	National AI Centre	The six practices behind the high-risk path and evidence pack
Privacy Act 1988 and the APPs	Australian Government	Personal information handling
ISO/IEC 42001	ISO and IEC	Manifest is certified; the assessment runs inside its certified AI management system
EU AI Act, Article 5	European Union	Prohibited practices, folded into red line 4

Glossary

Manifest

aOS	Manifest's Agentic Operating System: knowledge, agents and people on one governed layer
Agent	An AI system that plans and acts, such as reading a document and updating a ledger
AI register	A list of every AI system an organisation runs, with its owner and risk level
Evidence pack	The records Manifest hands a client to show the due diligence was done
FAIRA	Foundational AI Risk Assessment, the Queensland Government's AI risk framework
Human-in-the-loop	A person approves each decision before it takes effect
Red line	A use case Manifest will not build under any circumstances
Sovereign deployment	Data, models and audit trails run inside a jurisdiction and environment the client controls