The one-day evaluation

Can AI make evaluation quicker, cheaper, *and* better?

Julian King

October 2026

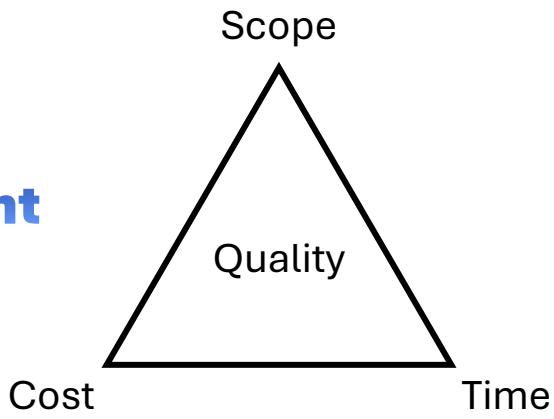
1

My use of AI

At work – a little	Off duty – a lot
Capacity building	Learning, exploring
Human-to-human(s)	Sense-making, writing
Advising people who use AI	Thought experiments…

2

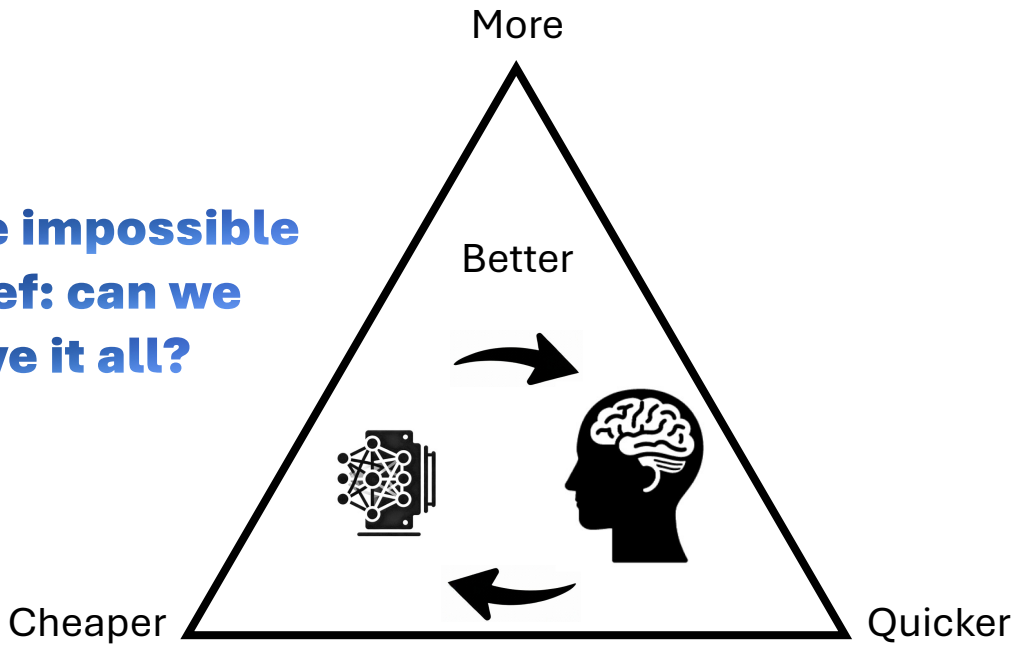
Project management triangle



julianking.co.nz

3

The impossible brief: can we have it all?



julianking.co.nz

4

Artificial Intelligence: a large and diverse toolbox

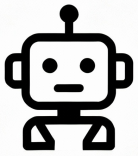
Large Language Model (LLM)
e.g., ChatGPT

Predictive/
forecasting models

Computer vision
and image processing

Statistical analysis

Speech and audio processing

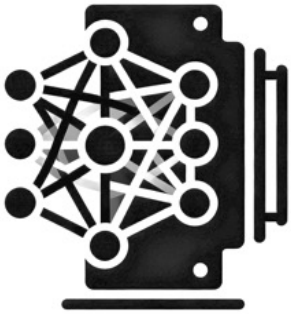


Agentic automation

julianking.co.nz

5

Pattern matching



LLM generates responses
from learned patterns

Understanding

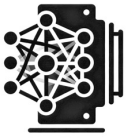
Interprets context,
deliberates about values,
and owns the judgement

julianking.co.nz

6

1-minute evaluation

50-day evaluation



One prompt, no iteration,
Inferred values,
no accountable evaluator

Stakeholder values,
reasoned judgements,
human accountability

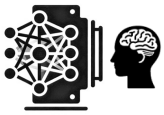
Cheap, but not fit for
consequential decisions

Expensive;
missed opportunities

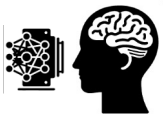
julianking.co.nz

7

Is there a middle ground?



1 day … 5 days … 10 days?



Evaluator and stakeholder-led, with AI assistance

How might we design for this?

julianking.co.nz

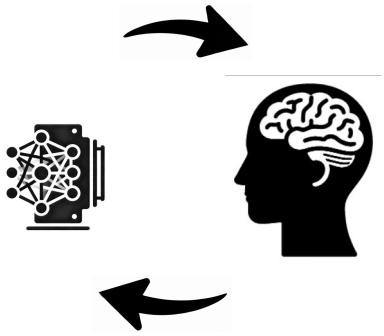
8

Comparative advantage

Each party specialises in what they do best

Traditionally applies to trade...

Can apply to evaluation roles...



julianking.co.nz

9

Comparative advantages

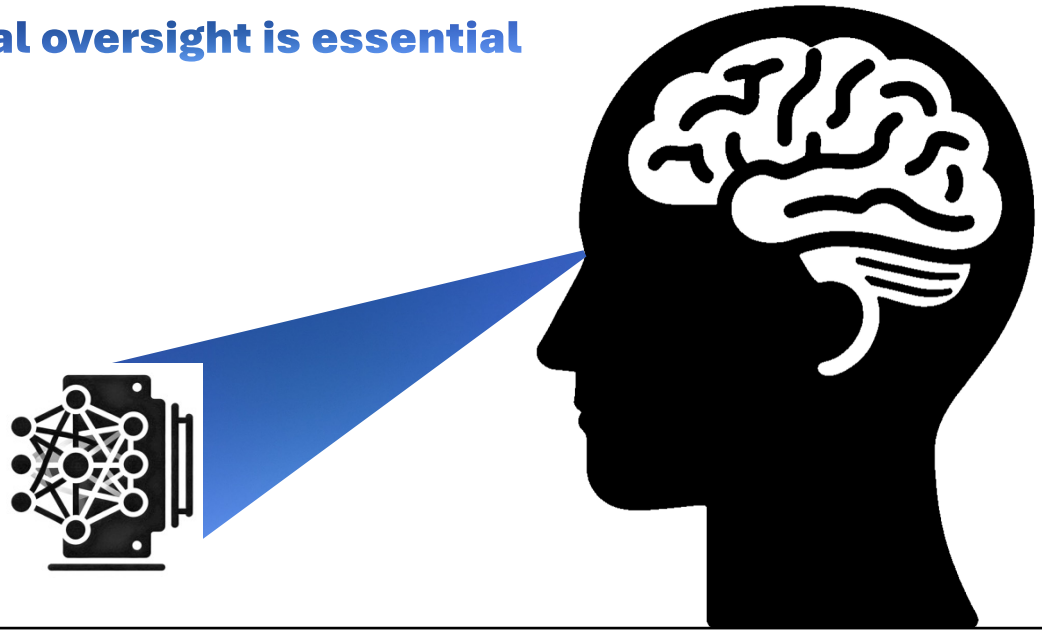
Helpful tools	Human responsibility
Fast, low-cost assistant	Evaluative judgement
Automating routine tasks	Relationships and trust
Brainstorming and scoping	Ethics and accountability
Rapid research and analysis	Creativity and adaptive thinking
Reviewing and stress-testing drafts	Curiosity and critical questioning
Prone to biases & confabulation	Prone to biases & heuristic shortcuts

julianking.co.nz

10

Humans remain accountable

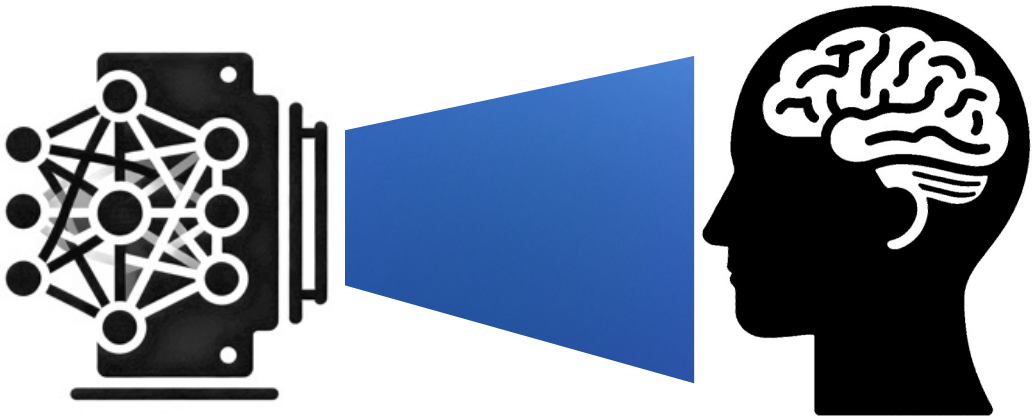
Ethical oversight is essential



julianking.co.nz

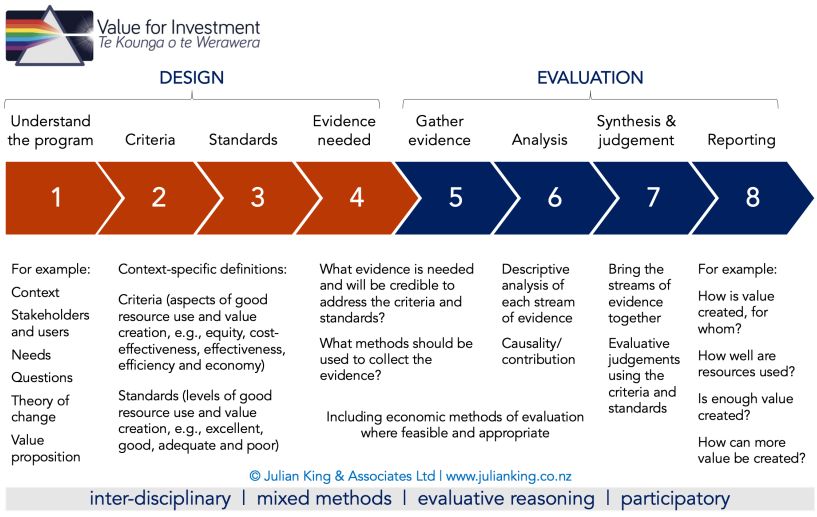
11

AI can challenge our thinking, too

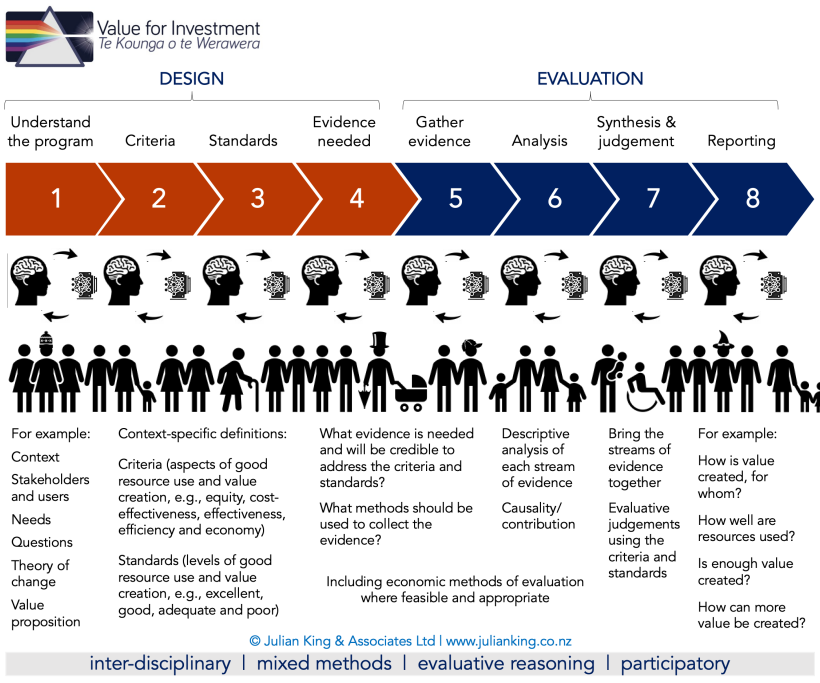


julianking.co.nz

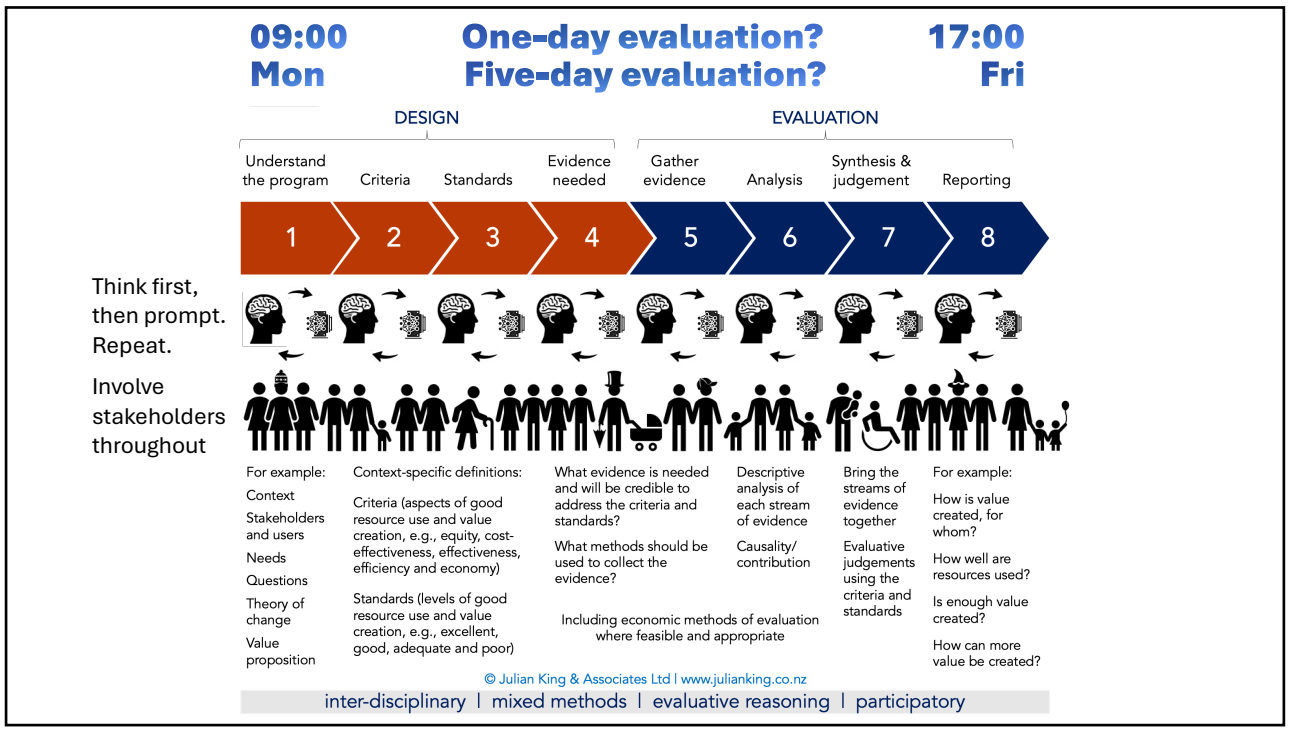
12


Value for Investment
Te Kounga o te Werawera
DESIGN
EVALUATION
Understand the program
Criteria
Standards
Evidence needed
Gather evidence
Analysis
Synthesis & judgement
Reporting
1
2
3
4
5
6
7
8
For example:
Context
Stakeholders and users
Needs
Questions
Theory of change
Value proposition
Context-specific definitions:
Criteria (aspects of good resource use and value creation, e.g., equity, cost-effectiveness, effectiveness, efficiency and economy)
Standards (levels of good resource use and value creation, e.g., excellent, good, adequate and poor)
What evidence is needed and will be credible to address the criteria and standards?
What methods should be used to collect the evidence?
Including economic methods of evaluation where feasible and appropriate
Descriptive analysis of each stream of evidence
Causality/ contribution
Bring the streams of evidence together
Evaluative judgements using the criteria and standards
For example:
How is value created, for whom?
How well are resources used?
Is enough value created?
How can more value be created?
© Julian King & Associates Ltd I www.julianking.co.nz
inter-disciplinary I mixed methods I evaluative reasoning I participatory

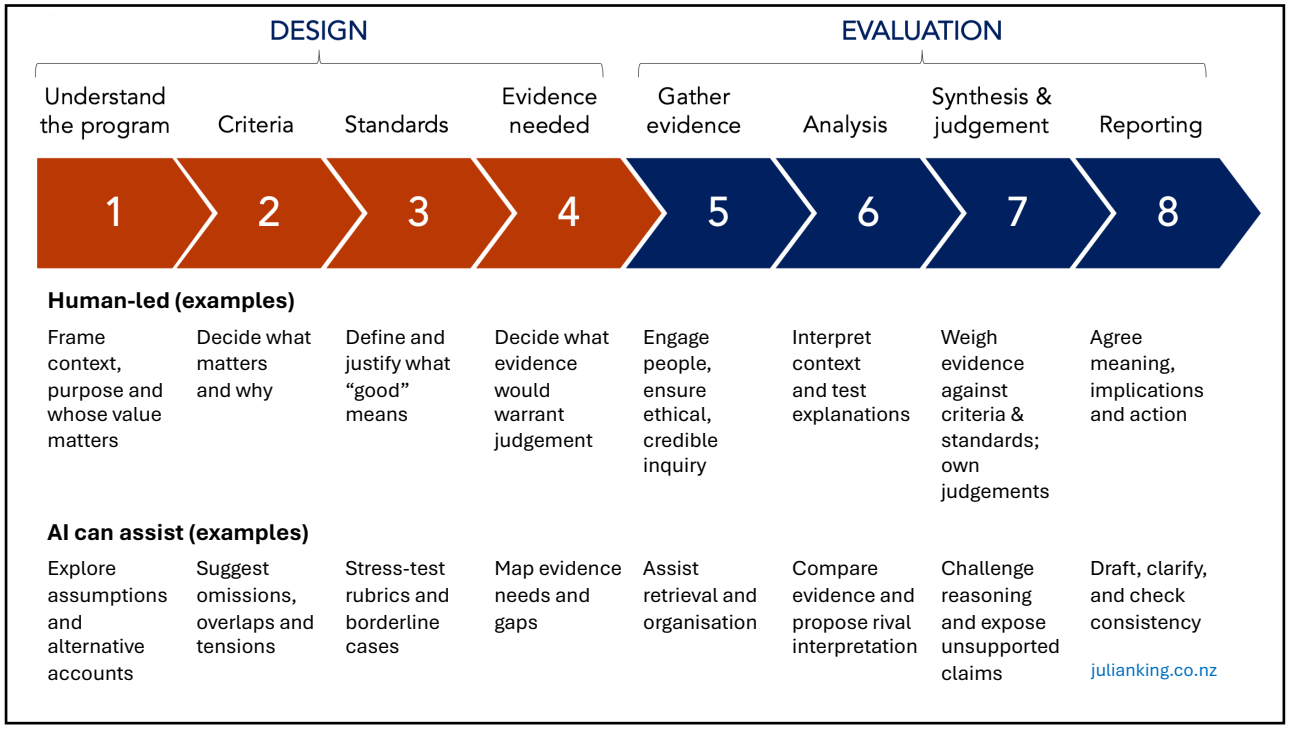
13


Value for Investment
Te Kounga o te Werawera
DESIGN
EVALUATION
Understand the program
Criteria
Standards
Evidence needed
Gather evidence
Analysis
Synthesis & judgement
Reporting
1
2
3
4
5
6
7
8
Think first, then prompt. Repeat.
Involve stakeholders throughout
For example:
Context
Stakeholders and users
Needs
Questions
Theory of change
Value proposition
Context-specific definitions:
Criteria (aspects of good resource use and value creation, e.g., equity, cost-effectiveness, effectiveness, efficiency and economy)
Standards (levels of good resource use and value creation, e.g., excellent, good, adequate and poor)
What evidence is needed and will be credible to address the criteria and standards?
What methods should be used to collect the evidence?
Including economic methods of evaluation where feasible and appropriate
Descriptive analysis of each stream of evidence
Causality/ contribution
Bring the streams of evidence together
Evaluative judgements using the criteria and standards
For example:
How is value created, for whom?
How well are resources used?
Is enough value created?
How can more value be created?
© Julian King & Associates Ltd I www.julianking.co.nz
inter-disciplinary I mixed methods I evaluative reasoning I participatory

14


09:00 Mon
One-day evaluation?
Five-day evaluation?
17:00 Fri
DESIGN
EVALUATION
Understand the program
Criteria
Standards
Evidence needed
Gather evidence
Analysis
Synthesis & judgement
Reporting
1
2
3
4
5
6
7
8
Think first, then prompt. Repeat.
Involve stakeholders throughout
For example:
Context
Stakeholders and users
Needs
Questions
Theory of change
Value proposition
Context-specific definitions:
Criteria (aspects of good resource use and value creation, e.g., equity, cost-effectiveness, effectiveness, efficiency and economy)
Standards (levels of good resource use and value creation, e.g., excellent, good, adequate and poor)
What evidence is needed and will be credible to address the criteria and standards?
What methods should be used to collect the evidence?
Including economic methods of evaluation where feasible and appropriate
Descriptive analysis of each stream of evidence
Causality/ contribution
Bring the streams of evidence together
Evaluative judgements using the criteria and standards
For example:
How is value created, for whom?
How well are resources used?
Is enough value created?
How can more value be created?
© Julian King & Associates Ltd I www.julianking.co.nz
inter-disciplinary I mixed methods I evaluative reasoning I participatory

15


DESIGN
EVALUATION
Understand the program
Criteria
Standards
Evidence needed
Gather evidence
Analysis
Synthesis & judgement
Reporting
1
2
3
4
5
6
7
8
Human-led (examples)
Frame context, purpose and whose value matters
Decide what matters and why
Define and justify what "good" means
Decide what evidence would warrant judgement
Engage people, ensure ethical, credible inquiry
Interpret context and test explanations
Weigh evidence against criteria & standards; own judgements
Agree meaning, implications and action
AI can assist (examples)
Explore assumptions and alternative accounts
Suggest omissions, overlaps and tensions
Stress-test rubrics and borderline cases
Map evidence needs and gaps
Assist retrieval and organisation
Compare evidence and propose rival interpretation
Challenge reasoning and expose unsupported claims
Draft, clarify, and check consistency
julianking.co.nz

16

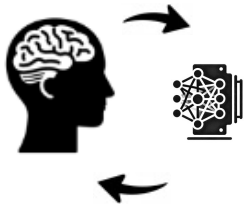
Can we design human-led, AI-assisted evaluation that is not just quicker and cheaper, but *better?*

This is one question – not the whole AI conversation.

Social and environmental risks are real;
Professional caution is appropriate.

AI is already here;
Used carefully, it may help us improve evaluation.

AI is evolving – ongoing conversation is needed.



17

Human-led, AI-enhanced evaluation

Value for Investment
Te Kounga o te Werawera

substack

18