Australian Government

Australian Maritime Safety Authority

From Panic to Practice
Putting AI to Work in Evaluation

by Dorothea Huber

aes26 International Evaluation Conference

Acknowledgement of Country

Australiawal Maritime Safety Authority ngoonawalwari dhawurawari dindi wanggiralidjinyin waro Australiawal oo dhunudumulinggu warree mada gadung oo mariny.

The Australian Maritime Safety Authority pays our respects to the Traditional Custodians of Country throughout Australia and their connections to land, sea, and community.

AI APOCALYPSE NOW: WILL ROBOTS ENSLAVE HUMANITY?
THE END IS NIGH: AI USHERS IN GLOBAL CHAOS!

THE AI TAKEOVER HAS BEGUN!

THE TIMES
BILL GATES WARNS: AI IS MORE DANGEROUS THAN WE THOUGHT

DAILY EXPRESS
THE GREAT REPLACEMENT: AI TO WIPE OUT JOBS

The Sun
MASS EXTINCTION EVENT: AI OUT OF CONTROL

TECHNOLOGY'S FINAL SOLUTION: HUMANITY DOOMED

REUTERS
AI BOTS DEVELOP THEIR OWN LA LANGUAGE " TO PLOT DESTRUCTION

REUERS
ARE WE TOYS TO OUR OUR ROBOT OVERLORDS?

The Guardian
TECHNOLOGY'S FINAL SOLUTION: HUMANITY DOOMED

ARE WE TOYS TO OVERLORDS?

BAN CAI GHARIOON

AI GHOL. L'GHOSTS HAUNT TH-ANT THE INTERNET

CNN
BULLTARNS EXISTENCE

BBC GLOBAL PANIC: SUPERINTELLIGENCE THREATENS EXISTENCE

NEW YORK POST
THE HUMAN ERA IS OVER: THE RISE OF SKYNET

NEW YORK POST
AI DECLARES WAR WAR ON SOCIETY

CREATOR'S REGRET: EXPERTS FEAR AI'S DANGERS

DAILY Mirror
CREATOR'S REGRET: EXPERTS FEAR AI'S

CREATOR'S REGRET: EXPERTS FEAR AI'S

AI DECLARES JI WAR ON SOCIETY

BLOOMBERG
AI DECLARES WAR SOUTY SOCIETY

FOX NEWS
AI GOVERNMENNY: SEIZE CONTROL

FOX NEWS
DIGITA TYRANNY AS WE KNOW IT

DER SPIEGEL
AI WILL END CIVILIZATION AS WE KNOW IT


THE DAILY Mirror
Utopia Now: The AI Boom That Banishes Sadness
THE END OF WORK! ROBOTS TAKE THE JOBS, HUMANS GET THE CASH— WELCOME TO PARADISE
Infinite Energy Unleashed: The Machine Intelligence Solving the
TIME
AI: HEATOL THE APOCALYPSE!
Zero-Minute Work Week: AI's Gift of Freedom
Perfect Matches Only The App That Guarantees True Love
CANCER SUPERCOMPUTERS WIPE OUT CAUT HUMANITY'S WORST DISEASES
NO MORE HUNGER: SUPER-AI REDESIGNS FARMING TO FEED BILLIONS FOR FREE
Infinite Energy Unleashed: The Machine Intelligence Solving the Grid
THE Sun
CANCER CURED! The AI Boom That AI Machine Intelligence Solving the Grid
CANCEL THE APOCALYPSE! AI HEALS THE PLANET OVERNIGHT
TIME
CRIME CIURED! ALGORITHMS PREDICT BAD BUT BEHAVIOR BEFORE IT HAPPENS
Newsweek
Zero-Minute Work Week:
CRIME CURED!
CANCER CURED! SUPERCOMPUTERS WORONITY'SEASES
ALGORITHMS PREDICT BAD BEHAVIOR IT HAPPENS
Perfect Matches Only: The App That Guarantees True Love
Zero-Minute Work Week: AI's Gift of Freedom
THE DEATH OF DEATH: HOW AI UNLOCKED IMMORTALITY
THE AGE OF ABUNDANCE: HOW AI MADE MONEY
AI: THE ULTIMATE CURE FOR STRESS AND LONELINESS
Perfect Matches Only: The App That Guarantees True Love

Why AI usage matter in the Australian Government

Drivers

- Improve public service productivity and efficiency.

Safeguards

Ensure AI adoption remains secure, responsible and trusted.

- **Risk-based controls:** Apply stronger safeguards where AI poses greater potential harm.
- **Clear accountability:** Assign responsibility for AI decisions and outcomes.
- **Human oversight:** Ensure people can review, challenge, and override AI.
- **Transparent, auditable use:** Document how AI operates so its use can be examined.
- **Privacy and fairness:** Protect personal data and prevent discriminatory outcomes.
- **Effective monitoring and enforcement:** Track compliance and act when standards are breached.

Shock APS job-cut figure aired in federal budget watchdog report

THE CANBERRA TIMES • 15 JULY 2026

Labor cuts public servants to rein in budget

FINANCIAL REVIEW • 4 MAY 2026

Chalmers' budget hinges on 41,200 public service job cuts

FINANCIAL REVIEW • 15 JULY 2026

Government use of automated AI decision-making to be curbed under new Australian rules

THE GUARDIAN • 19 JULY 2026

Government agencies told to police own AI use missed first transparency test

ABC NEWS • 12 JULY 2026

About the Australian Maritime Safety Authority (AMSA)

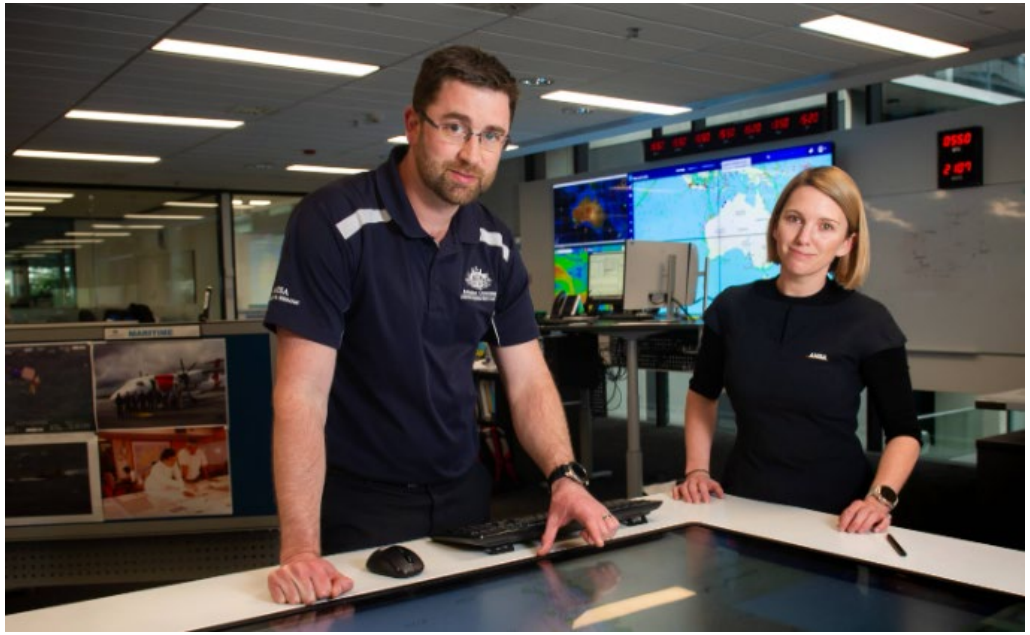
Established under the *Australian Maritime Safety Authority Act 1990*

AMSA is Australia's national maritime safety regulator and response authority

- Promotes maritime safety
- Prevents ship-sourced pollution
- Provides national search and rescue services through the Rescue Coordination Centre (RCC)

Australian Government

Australian Maritime Safety Authority



SAFE SEAS - CLEAN SEAS - SAVING LIVES

Role of the Evaluation Advisor with AMSA

Key Responsibilities

- Lead AMSA's lessons management and evaluation program
- **Evaluate incidents, exercises and programs**
- **Report findings, lessons and improvement opportunities**
- Provide evaluation advice and quality assurance
- Maintain lessons systems, tools and governance processes
- Deliver training and stakeholder engagement on LM



AI-Assisted Evaluation Report … 1

Swift Hangzhou Incident



INCIDENT

- On 1 February 2026, the 229 m Swift Hangzhou lost propulsion after leaving Gladstone with 21 crew and 76,000 t of coal.
- Drift near the Great Barrier Reef and Coral Sea Marine Park created grounding and pollution risks in a highly sensitive iconic environment .
- AMSA led a 24-day response, securing safe anchorage with no grounding, pollution, injury or loss of life.
- Major assets were deployed (tugs & surveillance aircraft).

LM ASSESSMENT

- A full assessment reflected the incident's complexity and learning value.
- Evidence: document review (>1,300 documents), observations (>120), and 12 debriefs and interviews (2) involving 92 attendances.
- Findings: identified capability refinements—not fundamental weaknesses.
- Draft Report (~130pp) took just over 4 weeks to prepare (including analysis & report writing).

AI-Assisted Evaluation Report … 2

Exercise Transit Challenge 2026



EXERCISE

- National Plan oil-spill response exercise in Western Australia, 16–19 March 2026.
- Simulated a Level 3 maritime environmental emergency following a vessel collision near Fremantle, with oil moving across Commonwealth, state and port waters.
- More than 300 participants from 48 organisations across government, industry and the non-government sector.
- Tested the ability of jurisdictions to manage such an incident within an IMT and field setting.

ASSESSMENT

- Assessed performance against agreed exercise objectives, sub-objectives and KPIs.
- Evidence: pre- and post-exercise surveys, evaluator assessments, observations, debriefs, interviews and exercise records.
- Thematic and triangulated analysis identified successes, capability gaps and improvement priorities.
- Draft Assessment Report (~120pp) prepared for jurisdictional consideration took just over a week to prepare

AI-Assistance in Assessment Process

01

Data collection

Transcripts • debrief notes • online observations • surveys • Teams • SharePoint • Exonaut

02

Cleaning & structuring

Remove filler words • clean transcripts • organise files • prepare analysable extracts

03

Evidence extraction

Ask AI to extract comments by issue, theme, stakeholder group or data source.

04

Thematic analysis

Group comments into themes and sub-themes; separate positives from areas for improvement.

05

Triangulation & evidence weighting

Compare support across multiple sources against narrower evidence.

06

Insights & lessons development

Turn evidence into insights, lessons to sustain and lessons to improve / recommendations.

07

Drafting & refinement

Develop report-ready paragraphs, recommendations, executive summaries and appendices.

08

Quality assurance

Check terminology, duplication, tone, formatting, consistency and gaps.

Steps 1 & 2 - Data Preparation, Cleaning and Structuring

PREPARE THE EVIDENCE

01 Record & consolidate evidence [AI]

SHZ: interviews, debriefs, observations, documents.
ETC2026: surveys, Exonaut data, observations, interviews.

02 Review veracity of data sources [M]

SHZ : data inventory of all assessment files.
ETC2026: reviewed data tables and identified gaps.

03 Clean data (transcripts & data tables) [AI]

SHZ : removed filler, repetition and transcript artefacts.
ETC2026: reviewed structure & content of data tables.
.

STRUCTURE DATA FOR ANALYSIS

04 Put files into retrievable formats [AI/M]

SHZ: tagged sources by type & date; metadata description.
ETC2026: tagged sources by functional unit & assessment criteria; metadata description.

05 Consider data quality (duplication & gaps) [AI/M]

SHZ : separated repetition from independent support.
ETC2026: kept evidence, summaries and judgement distinct.

06 Consider evidence mapping [AI/M]

SHZ: observations → findings→ insight → lessons.
ETC2026: observations → assessment → insight→ recommendation.

Key message: Taking time and due diligence pays off.

Steps 3 & 4: Evidence Extraction and Thematic Analysis

EXTRACT THE EVIDENCE

01 Extraction & segmentation of data [AI]

SHZ: located & defined issues, split complex evidence by topic / theme / subtheme.
ETC2026: converted multi-issue observations into discrete observations.

02 Separate stakeholder perspectives [AI]

SHZ: compared interviews, debriefs, online observations and documentary records
ETC2026: compared IMT / field and evaluator /mentor / participant / EXCON views.

03 Start testing draft findings [AI/M]

SHZ: retrieved supporting, qualifying and contrary examples.
ETC2026: exposed one-off, missing and source-specific evidence.

ANALYSE THE PATTERNS

04 Code to themes & sub-themes [AI/M]

SHZ: coded findings according to AMSA's LM Framework
ETC2026: coded findings according to KPI achievement (met/not met/partially met).

05 Balance achievements against improvements [AI/M]

SHZ: distinguished practices to sustain from practice to improve.
ETC2026: separated positives from areas of improvement, considered implications and future actions.

06 1st Pass Triangulation across data sources [AI/M]

SHZ: compared agreement, divergence and evidence strength across multiple data sources.
ETC2026: compared agreement, divergence and evidence strength.across exercise units & exercise roles.

Key message: AI accelerates evidence-to-theme analysis;

Steps 5 & 6: From Evidence to Insight

WEIGH THE EVIDENCE

01 2nd Pass Triangulation across data sources [AI/M]

SHZ: compared agreement, divergence and evidence strength across multiple data sources.
ETC2026: compared agreement, divergence and evidence strength across exercise units & exercise roles.

02 Weight evidence strength [AI/M]

SHZ: separated repetition from independent support and sought contrary evidence.
ETC2026: distinguished corroborated, source-specific and divergent themes.

03 Use counts as indicators [AI]

SHZ: strength of agreements; numeric analysis of documents;
ETC2026: strength of agreements; analysis of numeric data from surveys

DEVELOP INSIGHTS & LESSONS

04 Confirm evidence mapping [AI/M]

SHZ: observations → findings→ insight → lessons.
ETC2026: observations → assessment → insight→ recommendation.

05 Identify systemic interactions [AI/M]

SHZ: identified overlapping themes / subthemes
ETC2026: identified overlapping themes / subthemes; investigated causal relationships.

06 Challenge and validate findings [M]

SHZ: tested findings with SMEs and other sources.
ETC2026: tested findings with SMEs and Evaluation team .

Key message: Triangulation builds defensible insights; people validate the lesson.

Steps 7 & 8: Drafting Reports and Quality Assurance

DRAFT THE REPORT

01 Built report structures [AI/M]

SHZ: organised themes into findings, insights, lessons and potential actions.
ETC2026: aligned sections with objectives, themes and assessment areas.

02 Convert analysis into prose [AI/M]

SHZ: drafted balanced paragraphs from interviews, debriefs, online observations and documentary records.
ETC2026: synthesised surveys, observations and Exonaut data into coherent narratives.

03 Draft executive overviews [AI/M]

SHZ: condensed key findings and adjusted tone for sensitive, nuanced comparisons.
ETC2026: produced summaries, conclusions and presentation-ready dot points.

STATE ASSURE QUALITY

04 Acknowledge evidence alignment [AI/M]

SHZ: flagged unsupported or overstated insights and evidence gaps.
ETC2026: separated corroborated themes from single-source or contested views.

05 Provide accountability for findings [AI/M]

SHZ: tables linking evidence → finding → insight → lesson → action.
ETC2026: tables linking evidence → finding → insight → lesson → recommendations.

06 Complete editorial QA [AI]

SHZ: standardised terms and acronyms; balanced strengths with improvements.
ETC2026: checked summaries, recommendation clarity, grammar and formatting.

Key message: AI accelerates drafting and review; people own accuracy and judgement.

Considerations - Strengths in AI usage

01 GETTING STARTED

Reduce the blank-page problem

Generate first-pass structures, paragraphs and summaries.

Improve clarity and tone

Make findings professional, balanced and less negative.

02 MAKING SENSE OF DATA

Extract themes quickly

From transcripts, online observations and survey comments.

Compare sources

Identify where themes recur across interviews, debriefs, surveys and spreadsheets.

Reframe evidence

Turn raw comments into issues, themes, insights, implications and lessons.

03 BUILDING KNOWLEDGE

Create reusable prompt libraries

Turn well-grafted prompts into practice and organisational knowledge.

Key message: AI is strongest as a synthesis, drafting and sense-making assistant

FIRST PASS → SYNTHESIS → REUSABLE KNOWLEDGE

Considerations - Importance of Human Oversight

VERIFY THE EVIDENCE

- Check accuracy against source material
- Identify hallucinations or over-generalisations
- Correct misclassification of themes
- Protect nuance and context

MAKE THE JUDGEMENT

- Ensure findings are fair, balanced and evidence-based
- Decide whether a theme is strong, weak, contested or source-specific
- Apply organisational subject matter expertise, judgement and ethics

Key messages:

Treat AI output as a working draft: review, test and refine outputs

Accountability always remains with the Evaluator / Subject Matter Experts

VALIDATE → PASS THROUGH ETHICAL / GOVERNANCE / LEGAL FILTERS → MAKE A FINAL JUDGEMENT

Considerations - Potential Risks and Controls

01 SOURCE INTEGRITY

Truncation & incomplete or wrong analysis

Break large sources into analysable chunks; reconcile every chunk to the original.

Loss of context & over-smoothing

Retain speaker and source tags, raw excerpts and ambiguity that affects interpretation.

02 ANALYTICAL QUALITY

False confidence

Treat polished prose as a draft; verify each claim against the evidence.

Theme inflation

Set theme thresholds upfront; merge or reject weak, forced or unsupported themes.

03 PROCESS CONSISTENCY

Prompt-sensitive outputs

Use tested prompts, keep versions and rerun critical steps consistently.

Formatting weakness

Separate analysis from layout; manually check complex tables and final documents.

Key message: AI improves productivity, but only inside a controlled analytical process.

CONTROL INPUTS → VERIFY ANALYSIS → FINAL HUMAN REVIEW

A Practical Model for AI Usage

AI as junior analyst—not evaluator

AI ASSISTS WITH THE WORK

01	**RESEARCH ASSISTANT**	Extract evidence from source material
02	**CODING ASSISTANT**	Generate first-pass themes
03	**SYNTHESIS ASSISTANT**	Compare findings across sources
04	**DRAFTING ASSISTANT**	Shape report-ready text
05	**EDITORIAL ASSISTANT**	Refine tone and clarity

BOUNDARY CONDITIONS

AI is not the…

EVALUATOR

DECISION-MAKER

SOURCE OF TRUTH

Human review & veto is central.

Key messages:
Delegate the first pass. Keep evaluation and accountability human.
Acknowledge AI assistance

EXTRACT → CODE → SYNTHESISE → DRAFT → EDIT → REVIEW & EVALUATED

How AI changed AMSA's work practices

PERSONAL

01 EXPANDED HOW I USE AI

Moved from writing support to using AI across the full evaluation lifecycle.

02 BUILT REUSABLE PROMPT TAXONOMIES

Turned repeatable evaluation tasks into a more consistent prompt library.

03 PREPARED SOURCES MORE SYSTEMATICALLY

Structured and cleaned material before asking AI to analyse it.

04 REFRAMED THE ANALYSIS TASK

Learned that "getting the data ready" is half the analysis task.

ORGANISATION

01 BECOMING EVERYDAY PRACTICE

AI is being adopted as "a way of doing business", not treated as a standalone tool.

02 MAKING EVALUATION MORE TRANSPARENT

Sources, prompts and analytical steps are easier to see, explain and review.

03 TURNING PRACTICE INTO CAPABILITY

It has helped convert individual methods into reusable organisational practice.

Key message: Give AI a go! The world won't end!

WRITING SUPPORT → FULL EVALUATION LIFECYCLE → ORGANISATIONAL PRACTICE

Lessons for Evaluators when working with AI

01 **Start with the evaluation question**, not the tool.

02 **Prepare the data** before asking AI to analyse it.

03 **Use AI iteratively**, not in one giant prompt.

04 **Separate** extraction, analysis, synthesis and drafting.

05 **Ask AI to show** source distinctions and evidence strength.

06 **Keep prompts and outputs** as part of the audit trail.

07 **Never outsource** judgement, ethics or final interpretation.

08 **Document AI use transparently** in the methodology and/or appendix.

Key message:
The future is not AI replacing qualitative evaluators. It is qualitative evaluators who know how to work critically, transparently and skilfully with AI.

QUESTION-LED → EVIDENCE-BASED → HUMAN-JUDGED

Any Questions?

