STRATEGIC ANALYSIS

The evolution of human-centred design in the age of AI

Implications for research, service design and user experience in government

Service Design

July 2026 | Final draft

This document follows the conventions of the Australian Government Style Manual (stylemanual.gov.au).

Contents

Executive summary

Artificial intelligence (AI) is not merely a new technology capability. It represents a fundamental shift in how people interact with digital systems and how organisations design services.

For decades, human-centred design (HCD) has focused on creating understandable, predictable and usable interfaces. Users interacted with structured systems through buttons, menus, forms and predefined workflows. Designers could largely anticipate user journeys because systems behaved in deterministic ways.

Generative AI fundamentally changes this relationship. Instead of navigating systems through interfaces, users increasingly interact through conversations, prompts and natural language. The experience is no longer entirely designed in advance. AI systems generate responses dynamically, creating interactions that are probabilistic rather than deterministic. Two users may ask the same question and receive different answers. A single user may receive different answers to the same question on different occasions.

As a result, human-centred design is moving beyond the design of screens and workflows towards the design of human–AI relationships. This shift has significant implications for service design, user research, usability testing and organisational capability.

This shift is now formally recognised across Australian government policy. The Digital Transformation Agency's updated Policy for the responsible use of AI in government (v2.0, effective 15 December 2025) requires agencies to maintain registers of AI use cases, assign accountable owners, complete AI impact assessments and publish transparency statements written in clear, plain language (DTA 2025a). In NSW, the refreshed AI Assessment Framework (January 2026) mandates risk-based assessment of every AI use case across its full lifecycle, with public-facing AI chatbots assessed as high risk precisely because they can influence the advice people receive (Digital NSW 2026). These obligations cannot be met by technologists alone. They demand exactly the capabilities that human-centred design practitioners hold: understanding user needs, testing trust and comprehension, designing for transparency, and evaluating services against human outcomes.

The evidence from early government deployments reinforces this. The UK Government Digital Service ran two public pilots of GOV.UK Chat involving more than 10,000 users and 26,000 questions — one of the largest user research exercises GDS has ever conducted — before opening access more widely (GDS 2026). Conversely, the Royal Commission into the Robodebt Scheme stands as a lasting reminder of what happens when automated systems are deployed without transparency, explainability or meaningful human oversight (Royal Commission 2023).

Key findings

- AI changes what designers design. The focus expands from navigation, screens and workflows to prompting behaviour, trust, transparency, explainability, confidence calibration, human oversight and error recovery.
- AI changes the nature of research. Transcription, coding, clustering and initial synthesis can be accelerated by AI, shifting the researcher's value from producing insights to interpreting them — understanding context, nuance, motivation, contradiction and organisational implication.

- AI introduces new user experience risks that traditional usability practice does not address: hallucinations, automation bias, over-trust and lack of explainability.
- Traditional usability testing is no longer sufficient. Task completion remains necessary but must be supplemented by testing for trust, explainability, error recovery, confidence calibration and user adaptation.
- Human-centred design becomes more important, not less. As systems become more autonomous and less predictable, the disciplines of research, behaviour, ethics, accessibility, inclusion and service design become critical organisational capabilities — a position now embedded in Commonwealth and state AI policy.

Key proposition

AI is not replacing human-centred design. It is transforming what human-centred design must design.

The focus of HCD is expanding from creating usable interfaces to shaping trustworthy, transparent and collaborative relationships between people and intelligent systems. Organisations that recognise this shift early will be better positioned to design services that are both technologically advanced and genuinely human-centred — and to meet their obligations under the Commonwealth and NSW AI policy frameworks.

Why AI changes human-centred design

Human-centred design matured in an era of deterministic systems. ISO 9241-210 defines HCD as an approach that makes systems usable and useful by focusing on users, their needs and requirements. In practice, this meant designers could specify, in advance, every state a system could reach and every path a user could take. Research validated those paths; usability testing confirmed users could complete them.

Generative AI breaks three assumptions that underpin this model.

From deterministic to probabilistic systems

Traditional software produces the same output for the same input every time. Generative AI produces outputs drawn from a probability distribution. This 'generative variance' means designers can no longer guarantee what users will see. Design effort shifts from specifying outputs to shaping the conditions under which outputs are produced: system prompts, guardrails, grounding sources, confidence signals and recovery mechanisms. Users must be able to safely explore, reverse actions and give feedback when the AI gets it wrong.

From bounded to open-ended interaction

A form has a finite set of fields; a conversation does not. When the primary interaction is natural language, the 'interface' is effectively unbounded. Users will ask questions the team never anticipated, in phrasings no journey map predicted. During the GOV.UK Chat pilots, users asked 26,000 questions across tax, benefits and visas — a scale and variety of intent that no information architecture exercise could have fully specified in advance (GDS 2026).

From usability to trust as the core design problem

When systems were predictable, the central question was 'can the user complete the task?'. When systems are probabilistic, the central question becomes 'does the user trust this output appropriately — and can they tell when they shouldn't?'. Appropriate trust is now a designed property. It must be calibrated: too little trust and the service delivers no value; too much trust and users act on incorrect information. This is why the Australian Government's AI transparency standard requires agencies to explain their AI use in plain language and to disclose when the public interacts with AI without a human intermediary (DTA 2025a).

None of this makes the fundamentals of HCD obsolete. Accessibility, inclusion, plain language and evidence-based iteration matter more than ever. What changes is the object of design: not the screen, but the relationship.

The shift from interface design to conversation design

Historically, designers focused on navigation, information architecture, screen layouts, forms, transactions and user journeys. Increasingly, they must also design prompting behaviour, trust cues, transparency mechanisms, explainability layers, confidence calibration, human oversight points and error recovery paths.

The table below summarises the shift.

Traditional HCD focus	Emerging HCD focus with AI
Navigation and information architecture	Conversation flows, prompt scaffolding and intent recognition
Screen layouts and visual hierarchy	Response formats, confidence signals and source attribution
Forms and predefined workflows	Open-ended dialogue with graceful constraint and redirection
Predictable user journeys	Probabilistic journeys with designed recovery and reversal paths
Error messages	Hallucination detection cues, correction affordances and feedback loops
Task completion	Appropriate trust, comprehension and human oversight

Designing the 'trust layer'

Explainable AI (XAI) in a service context does not mean exposing model internals. It means giving users enough insight into why the system produced a result that they can judge whether to rely on it — without overwhelming them with technical jargon. GOV.UK Chat operationalised this by placing a 'check this answer' link beneath every response so users can verify the source guidance, and by explaining the risk of inaccurate answers during onboarding (GDS 2024b). The GDS team later refined onboarding to three plain messages: state the purpose and scope of the tool, encourage users to double-check answers, and explain how conversation data is used (GDS 2026). This is conversation design as trust design.

The NSW Government applies the same logic at the assurance level. Under the refreshed AI Assessment Framework, a public-facing chatbot that helps people apply for services is automatically treated as high risk because it collects sensitive personal information and can influence the advice people receive, triggering privacy impact assessment, accessibility requirements and independent review (NSW Government 2026).

The new role of researchers

AI is already reducing many of the administrative and analytical burdens historically associated with research. Transcription, thematic coding, affinity clustering, initial synthesis and draft reporting can increasingly be accelerated through AI-assisted tools. This creates opportunities for research teams to process larger volumes of information and identify patterns more rapidly than manual approaches allow.

However, the value of research shifts away from producing insights and towards interpreting them. Researchers become less focused on documenting user behaviour and more focused on understanding context, nuance, motivations, contradictions and organisational implications. AI can identify patterns, but it cannot independently determine which patterns matter, which are meaningful, or which are likely to drive successful service outcomes.

The GOV.UK Chat evaluation illustrates what this hybrid model looks like in practice. The team combined automated evaluation — quantifying factual precision, recall, relevancy and groundedness — with structured manual evaluation of answer accuracy and interaction quality, red teaming with adversarial inputs, task-based benchmarking (99 remote usability tests), a diary study with 30 hours of interviews, survey analysis of 576 responses, and qualitative analysis of individual conversations (GDS 2026; PublicTechnology 2025). Machines handled scale; humans handled meaning.

Three shifts in practice

- **From notetaker to strategic director.** With AI drafting the first pass of synthesis, researchers direct inquiry: framing the questions that matter, challenging AI-generated themes, and connecting findings to service and policy decisions.
- **From data collection to data judgement.** AI-assisted analysis can amplify bias in the underlying data and can hallucinate themes that sound plausible. Researchers must audit AI outputs with the same rigour they would apply to a junior analyst's work — verifying quotes against transcripts and testing whether themes survive contact with raw evidence.
- **The 'human' gap remains.** AI identifies trends but lacks genuine empathy and contextual reasoning. Human-guided observation remains essential to understand user intent, emotional response and the unspoken context that shapes behaviour. This is why moderated, human-led testing remains irreplaceable for complex conversations and nuanced feedback.

Governance note for researchers using AI

Under the Commonwealth policy, foundational AI training is mandatory for all APS staff, and agencies must maintain an internal register of in-scope AI use cases with an accountable owner for each (DTA 2025a). Research teams using AI tools for analysis of user data should treat this as an in-scope use, apply the AI impact assessment tool, and follow their agency's privacy and data-handling obligations. In NSW, the AI Assessment Framework applies across the full lifecycle of any AI use, including internal research tooling (Digital NSW 2026).

New user experience risks

AI introduces user experience risks that have not traditionally existed in digital products. These risks mean traditional usability principles alone are no longer sufficient. Future design practice must explicitly address trust, transparency, recoverability and human judgement.

Hallucinations

AI systems can generate plausible but incorrect information. Unlike conventional software defects, hallucinations may appear credible and therefore be difficult for users to detect. The GOV.UK Chat experiment found cases where the system generated responses containing incorrect information presented as fact, mostly in response to ambiguous queries; GDS noted that, industry-wide, no one has reached 100 per cent accuracy in generated answers, and treated managing hallucination risk as intrinsic to working with the technology (GDS 2024a; GDS 2024b). Even after two years of iteration

and a shift to more capable models, GDS states that 'with any AI tool there is a risk of inaccurate responses' — hallucinations are reduced, not eliminated (PublicTechnology 2025). Independent testing of the launched product still identified misleading answers on complex tax questions (THINK Digital Partners 2026). Design implication: hallucination is a permanent design constraint, not a temporary bug. Services must make verification easy and correction obvious.

Automation bias

Users may place excessive trust in AI-generated recommendations simply because the output appears authoritative. This reduces critical thinking and increases the risk of poor decision-making. Robodebt is Australia's defining case study: the Royal Commission found the scheme was 'a crude and cruel mechanism, neither fair nor legal', and evidence showed that many people inside the system recognised its flaws but failed to challenge the outputs of the automated process (Royal Commission 2023; Monash University 2023). Automation bias operates on staff as much as on citizens. Design implication: services must build in structured prompts for human challenge, not just a nominal 'human in the loop'.

Over-trust and false confidence

Users may assume AI outputs are more accurate than they really are, particularly when systems provide fluent, highly persuasive responses. Fluency is not accuracy, but users routinely read it as such. Emerging research goes further, showing that conversational AI's agreeable, validating style can reinforce users' own inaccurate beliefs over repeated interactions (University of Exeter 2026). Design implication: confidence must be calibrated in the interface — through source links, uncertainty language and deliberate friction at consequential moments.

Lack of explainability

Users often struggle to understand how an AI arrived at a conclusion, reducing trust and limiting accountability. The Robodebt Royal Commission found that automated debt letters provided no explanation of the basis for the debts raised, meaning people could not assess whether a debt was incorrect or unlawful, and therefore could not effectively challenge it (UTS Human Technology Institute 2023). Procedural justice research confirms people more readily accept decisions as fair when they can see and understand the reasoning behind them (Monash University 2023). Design implication: explainability is not a technical nicety; it is the precondition for contestability, which the Commonwealth's AI Ethics Principles list as a core requirement (DISR 2019).

Case study: Robodebt — the cost of ignoring these risks

The Robodebt scheme (2015–2019) used income averaging to automatically raise welfare debts. It was opaque, provided no explanation to affected people, reversed the onus of proof, and lacked adequate oversight. The Royal Commission described it as a costly failure of public administration in both human and economic terms, and recommended legislative reform of automated decision-making, including clear review pathways and a body to monitor and audit ADM in government (Royal Commission 2023).

Robodebt predates generative AI — it was a simple automated system. Its lesson for AI-era service design is therefore stronger, not weaker: if a simple, rules-based system can cause large-scale harm when transparency, explainability and human oversight are absent, probabilistic AI systems demand deliberately stronger human-centred safeguards. This is the explicit rationale behind the assurance frameworks now mandated across the Commonwealth and NSW.

New usability testing methods for AI-based applications

Conventional usability testing asks whether users can complete tasks, locate information and navigate effectively. These questions remain necessary but are no longer sufficient. Because AI applications are unpredictable, task-completion metrics alone cannot tell us whether a service is safe, trusted or genuinely useful. Testing must expand to cover trust, system transparency and recovery.

Five additional evaluation dimensions

Dimension	What to test and how
Trust	Do users trust the system appropriately? Ask participants to rate confidence in specific answers, then compare against actual accuracy. Look for both under-trust (abandoning a correct answer) and over-trust (acting on a wrong one).
Explainability	Can users understand why the AI produced a result? Probe: 'Do you feel confident about why the AI provided this answer?' Observe whether users notice and use source links or explanations.
Error recovery	Can users recognise and correct AI mistakes? Intentionally seed sessions with known-incorrect or low-quality outputs and observe whether users identify the mistake, correct the AI, or reverse the action.
Confidence calibration	Can users distinguish high-confidence from low-confidence outputs? Test whether uncertainty cues (hedging language, confidence indicators, source quality) change user behaviour.
User adaptation	Do users learn to prompt more effectively over time? Longitudinal methods — diary studies, repeated sessions — reveal whether the human–AI relationship improves with use.

Method guidance

- **Test system capability against user intent.** Instead of asking users to find a specific button, give them a goal and observe how they prompt the AI to achieve it — and whether they understand how the AI arrived at its output.

- **Deliberately provoke failure.** Red teaming belongs in UX research, not just security. GDS systematically probed GOV.UK Chat with adversarial and edge-case inputs to uncover vulnerabilities and safety risks alongside its user testing (PublicTechnology 2025).
- **Use hybrid evaluation.** Use automated tools for preliminary accuracy, groundedness and accessibility audits at scale; rely on moderated, human-led testing for complex conversations, emotional responses and nuanced feedback. The GOV.UK Chat evaluation combined automated scoring of factual precision and groundedness with manual review of 419 conversations and 1,084 question–answer pairs (GDS 2026).
- **Test with subject-matter experts as well as users.** GDS worked with HMRC subject-matter experts to score the accuracy of answers against exemplar answers written by content designers (GDS 2024b). In an Australian context, this means pairing usability testing with policy and legal review of AI outputs.
- **Benchmark against the non-AI alternative.** GDS found Chat was faster and easier than browsing GOV.UK and comparable to search (PublicTechnology 2025). Both the Commonwealth AI impact assessment tool and the NSW AIAF require agencies to justify AI over non-AI alternatives — usability benchmarking provides that evidence (DTA 2025b; Digital NSW 2026).

Case study: GOV.UK Chat — testing trust at scale

Over 18 months, the UK Government Digital Service ran two public pilots of GOV.UK Chat in which more than 10,000 users asked 26,000 questions about government services. The team measured accuracy, reliability, speed, safety and user trust using task-based benchmarking (99 timed usability tests), a one-week diary study with 30 hours of interviews, 576 surveys, manual evaluation of over 1,000 question–answer pairs, and performance analysis of 4,764 user journeys. GDS describes it as one of the largest user research exercises it has ever conducted and the UK Government's biggest public test of generative AI (GDS 2026).

The design decisions that emerged — 'check this answer' source links under every response, a three-message onboarding covering purpose, verification and data use, and explicit expectation-setting about accuracy — are directly transferable patterns for Australian government services.

Case study: NSW EduChat — assurance as a design input

EduChat, the NSW Department of Education's AI chatbot for schools, used the NSW AI assurance process to identify high risks early — particularly around ethics, data quality and safe handling of sensitive topics such as mental health — and to define mitigation and live monitoring strategies before and after release. Digital NSW cites it as an exemplar of the assessment framework shaping a safer product, not just documenting one (Digital NSW 2024). For service designers, the lesson is that assurance frameworks are most valuable when treated as design research instruments rather than compliance paperwork.

New survey methods for AI-based applications

Traditional satisfaction surveys applied to AI services often yield 'hallucination-blind' responses: a user who confidently received wrong information will report high satisfaction. Surveys must therefore be structured around user perception, value, transparency, trust and safety — and triangulated with accuracy data from evaluation.

Value and relevance

- 'How accurately did the AI-generated recommendation match your intent for this task?' (scale 1–5)
- 'How much time did this AI feature save you compared to doing it manually?' (time bands, with 'it took longer' as an option)

Transparency and explainability

- 'When the AI provided its output, did you understand why it gave you that specific result?' (Yes / Somewhat / No / I didn't think about it)
- 'Was it easy to work out how to correct the AI when it made a mistake?' (Yes / No, with open text)

Trust and control

- 'How safe did you feel adjusting the settings or preferences that guide the AI's behaviour?' (scale 1–5)
- 'How often do you double-check the AI's output against other sources?' (Always / Often / Rarely / Never)

Design notes for AI surveys

- **Pair perception with behaviour.** The 'double-checking' question is a proxy for calibrated trust. Compare survey answers with observed verification behaviour (for example, click-through on source links) — GDS analysed both surveys and the actual questions respondents asked (GDS 2026).
- **Ask about the worst moment, not just the average.** Include a critical-incident item: 'Did the AI ever give you an answer you later found to be wrong? What happened next?' This surfaces hallucination impacts that satisfaction scales hide.
- **Segment by stakes.** Trust in an AI answer about opening hours is not comparable to trust in an answer about benefit eligibility. Tag survey responses by task consequence, consistent with the risk-based approach of the Commonwealth impact assessment tool and the NSW AIAF.
- **Use plain language.** The Style Manual's plain language guidance applies to research instruments too: avoid 'model', 'output' and 'parameters' in favour of 'the tool', 'the answer' and 'the settings'.

Implications for service design practice

For service design teams, AI represents both a capability opportunity and a methodological challenge. It also changes the policy environment in which service design operates: AI-enabled services now carry mandatory assurance, transparency and accountability obligations that sit naturally within — and elevate — service design practice.

A common misconception

It is sometimes assumed that AI automates human-centred design. The opposite may be true. As systems become more autonomous, unpredictable and complex, understanding human needs becomes increasingly critical. Organisations will need more expertise in research, behaviour, trust, ethics, accessibility, inclusion and service design to ensure AI solutions remain understandable, safe and valuable. The future role of HCD is not diminished by AI, but expanded.

What changes for teams

- Shift from designing screens to designing interactions and relationships between people and intelligent systems.
- Incorporate AI literacy into research and design practice. This is now a formal expectation: foundational AI training is mandatory for all APS staff under the updated Commonwealth policy (DTA 2025a).
- Develop new methods for testing trust, transparency and explainability, as set out in the previous sections.
- Build governance and ethics into design activities. The Commonwealth AI impact assessment tool is structured around Australia's AI Ethics Principles — including human-centred values, transparency and explainability, contestability and accountability — and explicitly requires identification of stakeholders, consideration of non-AI alternatives, and preparedness to intervene or disengage (DTA 2025b; DISR 2019). These are service design questions.
- Establish new approaches to evaluating AI-enabled services across their lifecycle, including live monitoring after release, in line with the NSW AIAF's lifecycle-based reassessment requirements (Digital NSW 2026).

The policy environment is converging on HCD

Three developments make this an opportune moment for service design to lead.

- **Commonwealth.** The Policy for the responsible use of AI in government v2.0 (effective 15 December 2025) and the APS AI Plan 2025 introduce use-case registers, accountable owners, mandatory impact assessment, Chief AI Officers, an AI Review Committee and mandatory transparency statements in plain language (DTA 2025a; DTA 2025c). Each mechanism requires evidence about human impact that service design and research are best placed to produce.
- **NSW.** The refreshed AI Assessment Framework (January 2026), developed with CSIRO's Data61 and aligned with the Commonwealth framework and the EU AI Act, reduces assessment time from days to under 30 minutes and automatically triggers safeguards — privacy impact assessment, accessibility requirements, AI Review Committee oversight — for high-risk uses such as public-facing chatbots. NSW also released an Australian-first guide to agentic AI in

October 2025, addressing systems that can plan and act, not just respond (NSW Government 2026; Digital NSW 2026).

- **Transparency and contestability.** The Office of the Australian Information Commissioner has highlighted that agencies have significant room to improve transparency of automated decision-making under FOI (OAIC 2026), and the Robodebt Royal Commission's recommendations continue to drive ADM reform. Explainable, contestable service experiences are becoming a legal and democratic expectation, not a design preference.

The practitioner of the future

The practitioner of the future is likely to combine human empathy, systems thinking, service design, behavioural insight and AI literacy to create experiences that blend machine capability with human judgement. In this environment, human-centred design remains the mechanism through which organisations ensure AI serves human needs rather than requiring humans to adapt to technology.

What service designers should start doing today

1. Complete foundational AI training and build hands-on fluency with generative AI tools. Credible design leadership on AI requires direct experience of how these systems succeed and fail.
2. Map your agency's AI obligations. Identify your accountable AI official, your agency's transparency statement, the use-case register, and where the AI impact assessment (Commonwealth) or AIAF (NSW) applies to work your team touches.
3. Embed service design into AI assurance. Volunteer research and design evidence — user needs, trust findings, accessibility results — into impact assessments, rather than letting them be completed as technical paperwork.
4. Extend your usability testing protocols now. Add trust, explainability, error recovery, confidence calibration and adaptation measures (section on testing methods) to your standard discussion guides before your first AI engagement, not during it.
5. Redesign survey instruments for AI services using the perception, transparency and trust structures in this paper, and pair them with behavioural data.
6. Adopt transferable patterns from proven deployments: source links under every AI answer, plain-language onboarding that sets expectations about accuracy, easy correction and reversal, and human escalation paths.
7. Pilot AI-assisted research on low-risk material. Trial AI transcription and first-pass synthesis on non-sensitive studies, with human verification, to build the hybrid research capability described in this paper.
8. Run a Robodebt lens over every AI concept. Ask early: can the affected person see why this happened, challenge it, and reach a human? If not, the design is not done.
9. Share findings through the Applied AI Community of Practice so that trust and usability evidence compounds across teams instead of being rediscovered project by project.

Future research agenda

This paper is a starting position. The following questions warrant structured internal research over the next 6 to 18 months, with findings shared to the Applied AI Community of Practice.

- Calibrated trust: what interface cues (source links, uncertainty language, confidence indicators) measurably improve users' ability to tell reliable from unreliable AI answers in our services?
- Error recovery in practice: when our users encounter an incorrect AI output, what do they actually do — and what design patterns raise correction rates?
- Equity and inclusion: how do conversational AI services perform for people with low digital literacy, low English proficiency, or who rely on assistive technology, and what does 'accessible AI' require beyond WCAG?
- Hybrid research quality: how does AI-assisted synthesis compare with human-only synthesis on our own studies, measured for accuracy, bias and insight quality?
- Longitudinal adaptation: how does user prompting behaviour and trust change over weeks of real use, and what should onboarding versus ongoing coaching each carry?
- Staff-facing automation bias: in internal AI tools, what conditions lead staff to accept incorrect AI outputs, and which challenge mechanisms work? (This directly addresses the Robodebt failure mode.)
- Assurance as design: how can AIAF / AI impact assessment evidence requirements be met through standard design research artefacts, so assurance and design are a single workflow?
- Agentic AI experiences: as agentic systems that plan and act enter service delivery, what new consent, oversight and intervention patterns do users need? (Building on the NSW agentic AI guidance, October 2025.)

References

References follow the author–date style recommended by the Australian Government Style Manual.

DISR (Department of Industry, Science and Resources) (2019) *Australia's AI Ethics Principles*, DISR, Australian Government, accessed July 2026. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles

DTA (Digital Transformation Agency) (2025a) *Policy for the responsible use of AI in government, version 2.0*, DTA, Australian Government, effective 15 December 2025. https://www.digital.gov.au/ai/ai-in-government-policy

DTA (Digital Transformation Agency) (2025b) *AI impact assessment tool and supporting guidance*, DTA, Australian Government. https://www.dta.gov.au

DTA (Digital Transformation Agency) (2025c) *AI Plan for the Australian Public Service 2025*, DTA, Department of Finance and APSC, Australian Government.

Digital NSW (2024) 'What's new in the Digital Assurance Framework and AI Assessment Framework — and why it matters', *Digital NSW*, Department of Customer Service, NSW Government. https://www.digital.nsw.gov.au

Digital NSW (2026) *NSW AI Assessment Framework*, Office for AI, Department of Customer Service, NSW Government, issued under Circular DCS-2024-04. https://www.digital.nsw.gov.au/policy/artificial-intelligence

GDS (Government Digital Service) (2024a) 'The findings of our first generative AI experiment: GOV.UK Chat', *Inside GOV.UK blog*, 18 January 2024. https://insidegovuk.blog.gov.uk

GDS (Government Digital Service) (2024b) 'We're running a private beta of GOV.UK Chat', *Inside GOV.UK blog*, 5 November 2024. https://insidegovuk.blog.gov.uk

GDS (Government Digital Service) (2026) '5 things we learned testing GOV.UK Chat: an AI assistant for government', *Inside GOV.UK blog*, 16 March 2026. https://insidegovuk.blog.gov.uk

Monash University (2023) 'Automation, uncertainty, and the Robodebt scheme', Monash University commentary on the Royal Commission.

NSW Government (2026) 'NSW strengthens AI oversight with modernised framework', media release, Minister for Customer Service and Digital Government, January 2026. https://www.nsw.gov.au

OAIC (Office of the Australian Information Commissioner) (2026) *Automated decision-making and public reporting under the Freedom of Information Act*, OAIC, Australian Government, 21 January 2026.

PublicTechnology (2025) 'GOV.UK Chat beating industry standards for accuracy', *PublicTechnology.net*, October 2025 (reporting GDS algorithmic transparency record).

Royal Commission into the Robodebt Scheme (2023) *Report of the Royal Commission into the Robodebt Scheme*, Commonwealth of Australia. https://robodebt.royalcommission.gov.au

THINK Digital Partners (2026) 'GOV.UK gets conversational as government launches AI chatbot', 18 May 2026.

University of Exeter (2026) 'Researchers say AI chatbots may blur the line between reality and delusion', summarising Osler L (2026) 'Hallucinating with AI: distributed delusions and "AI psychosis"', *Philosophy & Technology* 39(1).

UTS Human Technology Institute (2023) 'Reflections on Government Robodebt Royal Commission response', University of Technology Sydney.

Style reference: Australian Government Style Manual, https://www.stylemanual.gov.au