Digital NSW

AI agent usage and deployment guidance

NSW GOVERNMENT

Acknowledgement of Country

The NSW Department of Customer Service acknowledges the Traditional Custodians of the lands where we work and live. We celebrate the diversity of Aboriginal peoples and their ongoing cultures and connections to the lands and waters of NSW. We pay our respects to Elders past and present and acknowledge the Aboriginal and Torres Strait Islander people that contributed to the development of this document. We advise this resource may contain images, or names of deceased persons in photographs or historical content.

AI agent usage and deployment guidance
Published by the NSW Department of Customer Service.
First published: October 2025.

Overview and intended purpose

This guide helps NSW agencies know what to consider when developing policies, assurance processes and implementing AI agents. It also includes roles and responsibilities, use case screening, and checklists for pilots and production.

This is a guideline only. It is not mandatory and is not exhaustive. Agencies must still follow the frameworks and policies listed at the end of the document.

Understanding the role of AI agents

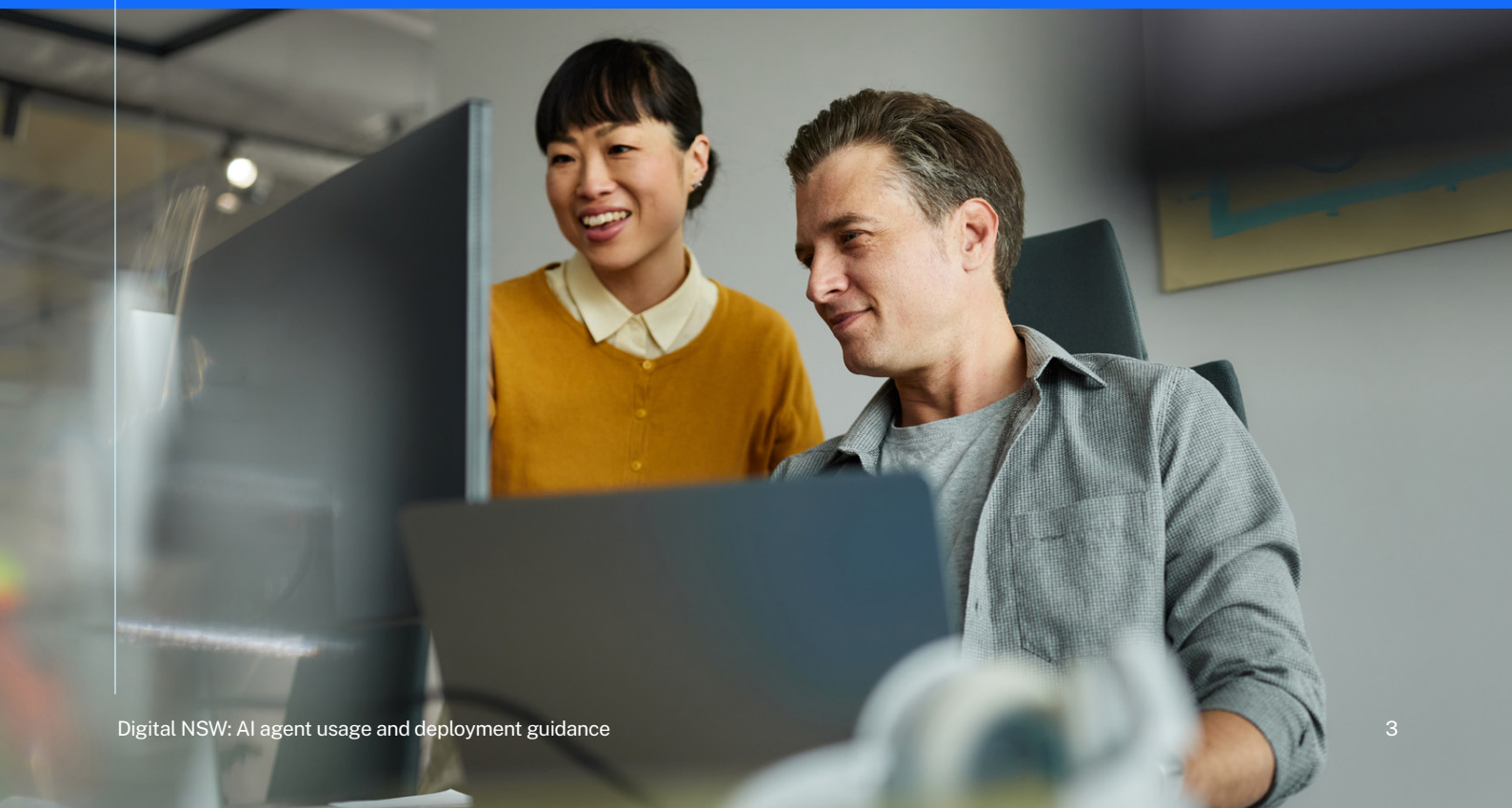
AI agents use models, tools, and data to identify tasks, make decisions, take action and learn over time to achieve goals. Because of their unique characteristics, AI agents must be well understood to ensure financial diligence, compliance and successful adoption – and to maximise benefits for government.

AI agent definitions vary. They offer different degrees of autonomy. Simple agents follow rules with limited actions, while other types plan and select tools independently to achieve objectives. This autonomy shapes the risk profile and sets agents apart from basic systems like chatbots that only generate text.

With agentic AI, the human's role shifts from completing tasks to managing AI agents that do some tasks for you. This shift calls for organisational agility, workforce retraining and a new focus on managing how multiple agents perform, interact and adapt.

AI agents are an emerging capability that introduce new risks in multi-agent deployments; for example, one agent can make a mistake that others copy or accept without question, creating false agreement (conformity bias). This is one of many considerations that underpin the need to always start with careful, controlled testing.

Agentic AI is evolving quickly. Please share your feedback to help expand and refine this guidance.



Summary of key points

Agentic AI is more than chat

AI agents can perceive, plan, reason, decide, learn and act autonomously with little or no human input to reach a goal.

Agents extend GenAI capabilities

Gen AI creates content. Agents act, learn, use tools, remember context, and work with other agents to achieve goals. Agentic AI is still in its infancy.

Assign ownership and guardrails up-front

Each agent needs a named owner, observability (including monitoring, audit logs), ideally a unique identity and clear escalation paths proportionate to risks.

Standards are still developing

Use this guide, share insights, proceed cautiously, and be prepared for updates to policy, governance and implementation as global standards evolve.

Considerations for an agency policy position

The following points are not mandatory for NSW Government agencies, but can help guide the development of agency-specific policies.

AI Agent ownership

Assign one Accountable Owner for each AI agent:

- **Business owner** – if the AI agent automates business or customer-facing processes.
- **IT owner** – if the AI agent automates system or infrastructure processes.

For agents that span business and technical functions or operate in multi-agent workflows, assign:

- a **Primary Accountable Owner** – typically from the business area
- a **Secondary Responsible Owner** – typically from IT, and
- a **System Owner** – for overseeing multi-agent workflows.

Responsibilities of an AI Agent Owner

Agent owner responsibilities should align with the level of risk associated with using the AI agent. Risk will vary between controlled testing environments and full production. Take a pragmatic approach to avoid unnecessary delays during experimentation.

AI Agent owner responsibilities include:

1. **Plan for accessibility and inclusion.** Design agents in consideration of cultural and linguistic differences and recognise when to defer to a human.
2. **Design for accountability.** Assign clear responsibility for each agent's system components, tasks, outputs and communications. Maintain transparency and oversight throughout the lifecycle, including decommissioning.
3. **Set and lock authority limits.** Define what the agent can do, what needs human approval, and what must remain human-only. Prevent agents from changing these limits.
4. **Monitor activity and maintain transparency.** Track real-time agent activity, tools usage, memory updates and costs. Record decisions and provide plain language explanations to promote transparency and surface anomalies.
5. **Disclose and label.** Inform people when they are interacting with an agent; explain its scope and watermark generated content where possible. Clearly state how user data will be used, in line with privacy obligations.
6. **Verify compliance, unwanted bias and quality continuously.** Use logs and dashboards to identify unwanted bias, drift, hallucinations, or other quality issues. Detect changes from self-learning behaviour and adjust safeguards. Include scheduled audits.
7. **Enforce data governance.** Apply privacy, security, and data rights controls across all datasets accessed or generated by agents.
8. **Maintain fail-safe plans.** Ensure you can detect, isolate, reverse or manually take control during outages or unexpected behaviour. Support this with an incident response plan.
9. **Manage upgrade dependencies.** Monitor changes to models, prompts, logic or system components that may affect agent behaviour or reliability.
10. **Benchmark human-AI synergy.** Compare human-only, AI-only and combined workflows to optimise collaboration. This helps test how effectively AI systems support human expertise in real-world situations.
11. **Enable safe innovation.** Use safe-to-fail environments and policy approaches to prototype low-risk use cases. Make sure agents support your business goals, reflect community values and meet public expectations.
12. **Improve Agent reliability.** Set service level agreements (SLAs) for outputs, validation, performance, escalation and failure thresholds. Define how ambiguity will be managed to improve reliability.
13. **Evaluate performance and value.** Review agent behaviour, user feedback and long-term performance and business impact. This helps confirm agents are reliable, fair and delivering value.
14. **Support continuous workforce training.** Help staff build the skills to think, collaborate and make decisions effectively with agents – and adapt to faster, more dynamic ways of working.
15. **Design for safe communication with other agents.** Define clear protocols, monitor interactions for errors or blind spots, and establish safeguards to prevent cascading failures.

Step 1. Use case screening checklist

Tick each box that applies. The more you tick, the stronger the case for an autonomous agent.

Checklist item	Use case need
☐ **Multi-step reasoning**	The workflow involves a sequence of decisions, where each result determines the next step.
☐ **Dynamic data sources**	Decisions rely on real-time inputs from changing application programming interfaces (APIs), databases or sensors – requiring instant responses.
☐ **Large-scale information retrieval**	The process must evaluate vast, fragmented data that's impractical to review manually.
☐ **Unstructured input**	Inputs arrive as free-text, speech, email, chat or documents, that need interpretation.
☐ **Complex or unpredictable paths**	The process changes based on user input, external events or real-world context.
☐ **Defined autonomy**	The process runs independently within clear boundaries, with little or no human intervention.
☐ **Proactive behaviour**	A need to ask clarifying questions, make suggestions or start follow-ups without prompts.
☐ **Clear, measurable KPIs**	Success is quantifiable (for example, cut cycle time by 50%). This lets you track the agent's return on investment.
☐ **High volume or high effort**	The task is frequent and labour-intensive; automation multiplies savings as scale grows.
☐ **Role-segmented collaboration**	The process involves multiple specialists, with built-in checks to maintain accuracy and quality.

If few boxes are ticked, consider simpler automation or a chatbot.

Step 2. Pilot readiness checklist

Piloting AI agents offers a valuable opportunity to foster a cyclic test-and-learn culture. Technically, an agent can be deployed in minutes. Embracing this speed will need adjustments to existing processes, mindsets and governance frameworks. Consider the following tasks before starting your AI agent pilot. These are not mandatory but highlight key readiness factors to consider.

Factor	Check or establish
☐ **Objectives and KPIs**	The agent's purpose, success metrics (for example, time saved), and exit criteria.
☐ **Prioritisation**	Prioritise low-risk, "safe-to-fail" pilots. Start with minimal autonomy and use human-in-the-loop to build trust.
☐ **Baseline Metrics**	Metrics to test ROI are documented, such as current process times, error rate and customer satisfaction score.
☐ **Vendor alignment**	The vendor product(s) aligns with your agency's values and risk controls.
☐ **Cost model**	Costs can be usage-based. Estimate early and set up tracking, alerts and controls to prevent budget overruns.
☐ **Data readiness**	Check that input and output data is accurate, relevant and protected. Close governance gaps that could impact the pilot.
☐ **Large language model (LLM) and tooling fit**	Document model choice and rationale. Use pre-approved LLMs, if available.
☐ **Architecture modularity**	List all tools, APIs, models and data components. Use open, widely adopted standards to avoid vendor lock-in.
☐ **Autonomy**	Define the agent's read, write and delete permissions, and where human checkpoints are needed. Higher autonomy can increase return on investment and risk.
☐ **Business continuity plan**	Define steps to manage outages or failures: what might cause them, how they'll be detected, and who does what.
☐ **Evaluations (Evals) of LLMs and agents**	Use a structured process to assess the agent and model's performance against expectations. Refine the solution based on evaluation results.
☐ **AgentOps pipeline**	Establish version control, configuration management, testing, deployment, monitoring and decommissioning to support safe, repeatable changes.
☐ **Traceability and contestability**	Ensure decision lineage is exportable (inputs → actions → outputs + timestamps). Essential if an output is challenged.
☐ **Knowledge retention**	Involve staff and document the agent's knowledge where needed to maintain essential skills and reduce vendor lock-in.
☐ **Guardrail effectiveness**	Validate that policy checks, anomaly detection and shutdown paths are effective across all agents to stop runaway behaviour.
☐ **Accountability reporting**	Ensure stakeholders can identify the agent's owner, department, and escalation contacts – with a process for changing ownership.
☐ **Multi-agent governance**	Define how multiple agents collaborate and manage shared risk. Plan how communication will be monitored for cascading errors, blind spots, and coordination breakdowns.

Step 3. Production readiness checklist

Ensure the pilot outcomes have been assessed and used to improve the solution and processes. Then, review the following tasks before production deployment.

Action	Why it matters
☐ **Obtain senior approval**	Binds executive ownership and accountability for production deployment.
☐ **Update the AIAF risk assessment**	Risks may change due to the scope of production deployment, self-learning, or changes to the tech environment, data sources or processes.
☐ **Activate guardrails**	Enforces compliance thresholds, cost limits, oversight and fail-safes. Include controls for contestability, right of reply, anomaly detection and unwanted bias.
☐ **Apply pilot lessons**	Confirms KPI gains and documented improvements are embedded before launch.
☐ **Register the agent in your asset register, and schedule reviews**	Maintains an authoritative inventory and lifecycle oversight.
☐ **Govern Data and LLMs**	Ensure data quality, privacy and authority. Confirms LLMs are adapted to agency needs and that agent actions, including data provenance and lineage, can be fully traced.
☐ **Validate multi-agent communications**	Confirms agents can safely coordinate, with controls to detect cascading errors, blind spots, and coordination breakdowns.
☐ **Confirm monitoring is active**	Set up dashboards and alerts to provide real-time visibility into agent performance and risks.
☐ **Audit tool access**	Check which tools, systems and instructions the agent can use. Confirm authorisation levels and apply appropriate safeguards.
☐ **Train the team**	Train the Agent Owner and support staff on risk controls, escalation paths and troubleshooting issues.
☐ **Test business continuity**	Make sure manual takeover and rollback steps work properly in a production-like environment.
☐ **Manage integrations**	Set integration limits for tools, data and connectors. Establish a process to review and re-approve any changes.
☐ **Check agent quota compliance**	Limit agents per owner (default 3–5), unless there's an approved exemption. This supports oversight and controlled scaling.

Comparison of AI and automated solutions

Choosing the wrong approach can waste budget, increase compliance risk, and reduce user trust. The following points are generalisations to help compare traditional automation, personal assistants, and AI agents. Use this guide to match the right solution to the task.

Use when...	Traditional automation (scripts / RPA / workflow)	Personal chat assistant	Autonomous AI agent
Task pattern	Steps are fixed, rule-based and repeatable.	Human-led tasks with faster drafting, coding or lookup.	Workflow branches or data shifts; the system must decide next steps.
Data volatility	Stable data; APIs rarely change.	Moderate; user reviews output.	High – live feeds, multiple sources, sensor or APIs.
Human role	Designs rules; reviews exceptions.	Remains “in the loop” for decisions and approvals.	Sets goals and guardrails; agent acts independently with oversight.
Governance load	Low; standard change-control suffices.	Medium; monitor prompts and outputs.	High; requires runtime monitoring, unwanted bias checks and shut-off triggers.
ROI driver	Reduces manual steps and licensing costs.	Boosts knowledge-worker productivity.	Automates complex, high-volume, dynamic workflows end-to-end.
Risk if misapplied	Breaks if rules change; staff must fix.	Gives wrong answers if not checked.	Incorrect actions or unexpected costs if guardrails fail.
Compounding intelligence	Static data repository.	GenAI improves through static re-training and use of past data.	Agents learn from real-world use; improve with experience.
Example	Invoice routing by fixed rules.	A legal AI co-pilot used by a subject matter expert.	Public enquiry agent → Retrieve data agent → Human → Response agent.

Note: Examples listed in this table are not exhaustive.

Feature comparison

Use this table as a general guide to compare different automation technologies and when to use them.

Feature	Traditional automation (scripts / RPA / workflow)	Personal chat assistant	Autonomous AI agent
Core capability	Follows fixed, rule-based steps.	Drafts, summarises, or searches to assist.	Plans and completes multi-step goals.
Trigger	Scheduled job, API or workflow event.	User prompt, click or voice command.	User initiated or System initiated – events, data changes, process steps.
User control & oversight	Human reviews exceptions only.	Human accepts or edits results.	Approval checkpoints optional; owner monitors anomalies.
Adaptability & learning	No learning – must be updated manually.	Learns from feedback; remains static until tuned.	Re-plans each run, storing new facts and strategies.
Governance load	Low; standard change control.	Medium; monitor prompts and outputs.	High; runtime monitoring, unwanted bias checks, shut-off triggers.
Deployment effort	Install script / bot; connect stable data.	Enable plug-in and check prompt controls.	Set up orchestration, guardrails and rollback processes.
Typical ROI driver	Removes repetitive manual tasks.	Boosts knowledge-worker productivity.	Automates complex, high-volume, dynamic workflows.
Data Context	Pre-defined, structured inputs	Uses static training data.	Live APIs, sensors, user feedback.
Guardrail Complexity	Simple validation rules.	Prompt or output filters.	Multi-layer limits and alerts.
Example	Invoice routing by static rules.	Writing customer emails in CRM.	Complaint intake → triage → case creation with live policy updates.

Further resources and guidance

The following is not a complete list but highlights key regulations and resources that support the safe, ethical and legal use of AI. Agencies should consult their governance and assurance teams to identify all relevant requirements for their business area.

Relevant legislation
Privacy and Personal Information Protection Act 1998 (PPIP Act) – NSW
Government Information (Public Access) Act 2009
State Records Act 1998
Public Interest Directions (PIDs) under the PPIP Act (exemptions)
Transport Administration Act 1988 (NSW)
Police Act 1990 (NSW)
Anti-Discrimination Act 1977 (NSW)
Australian Human Rights Commission Act 1986 (Cth)
Age Discrimination Act 2004 (Cth)
Disability Discrimination Act 1992 (Cth)
Racial Discrimination Act 1975 (Cth)
Sex Discrimination Act 1984 (Cth)

Further resources and guidance

Useful links

Artificial Intelligence | Digital NSW

www.digital.nsw.gov.au/policy/artificial-intelligence

Artificial Intelligence Ethics Policy | Digital NSW

www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy

NSW Artificial Intelligence Assessment Framework (AIAF) | Digital NSW

www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assessment-framework

Generative AI: basic guidance | Digital NSW

www.digital.nsw.gov.au/policy/artificial-intelligence/generative-ai-basic-guidance

Digital Assurance | Digital NSW

www.digital.nsw.gov.au/policy/digital-assurance

AI procurement essentials | info.buy.nsw

www.info.buy.nsw.gov.au

Privacy by Design (PbD) Fact Sheet | Information and Privacy Commission New South Wales

https://www.ipc.nsw.gov.au/resources/fact-sheet-privacy-design

Cyber Security NSW | Digital NSW

www.digital.nsw.gov.au/delivery/cyber-security

Australia's AI Ethics Principles | Australian Government

www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles

Glossary of key terms

The following definitions use plain language to support general guideline use. They are not intended as technical descriptions.

Term	Simple meaning
AI agent / Agentic AI	Software that identifies tasks to achieve a goal, makes decisions, acts, and adapts continuously through learning.
AI model	Software trained to find patterns, make predictions or understand data.
Agent Owner	Person responsible for the agent's actions, outputs and compliance.
AgentOps	Tools and processes used to monitor and manage how AI agents work in real time.
AIAF	The NSW Artificial Intelligence Assessment Framework (AIAF). It helps NSW agencies assess AI risk, apply controls and ensure compliance.
Autonomy	How much the agent can act without human approval or intervention.
Drift	When an AI system's behaviour or accuracy changes over time due to new data or context.
Guardrails	Limits that stop the agent from going outside safe limits.
Generative AI (GenAI)	AI that creates new content – such as text, images or code – based on what it has learned.
Hallucination	When AI outputs false or misleading information.
LLM (Large Language Model)	An AI model trained on large datasets to understand and generate human-like language.
Memory	Stored information that helps the agent remember actions, context, and decisions to improve over time.
Multi-agent system	A group of AI agents that work together on complex or connected tasks.

Digital NSW

NSW GOVERNMENT