Australian Government

National AI Centre

Understanding and using agentic AI

Helping organisations understand, assess and use AI agents safely.

October 2026

Contents

Introduction 4
How to use this resource 4
Definitions 4
Understand what agentic AI can do 5
What is agentic AI? 5
Example: How an AI agent might help 5
How to recognise agentic AI 5
Agentic AI can take different forms 6
Example: How multiple AI agents can work together 6
How can agentic AI help? 7
Examples of how AI agents could be used in business 8
Is agentic AI the right tool? 8
What job will it help with? 8
Who is responsible if something goes wrong? 9
Understand agentic AI risk 10
Common risks of agentic AI 10
What information will it use and store? 11
What actions can agentic AI take? 11
What access should an AI agent have? 12
Examples: matching access to the task 12
What can increase agentic AI risks? 12
When multiple AI agents interact 12
Three factors that can increase risk 13

Cyber security considerations 13
Use agentic AI safely 14
Use agentic AI on a small scale first 14
When to pause and get advice 15
Four safety checks before you turn it on 15
Why human oversight is not always enough 16
Manage unapproved use of agentic AI 17
Ask your vendor the right questions 17
Keep your AI agent under review 18
Before you use an agentic AI 18
Agentic AI checklist 19
1. Define the task 19
2. Understand the risks 19
3. Put controls in place 19
4. Prepare for deployment 19
5. Keep it under review 20

Introduction

> This resource provides practical information for Australian organisations about the use of agentic AI. It has been developed following consultation with experts in AI governance and adoption, and 175 representatives from across industry, academia, community and government. It draws on the latest safety research from the Australian AI Safety Institute and is aligned to international best practice AI governance.

How to use this resource

This resource is for small and medium-sized enterprises (SMEs) that are considering, buying, implementing or already using agentic AI systems.

It is designed for business owners, managers and staff who may not have specialist expertise but need to understand what agentic AI is, the risks it may create and what steps can be taken to manage common risks. This resource builds on the National AI Centre's *Guidance for AI Adoption* which sets out responsible practices for AI adoption.

Definitions

Agentic AI refers to a type of artificial intelligence that can plan, choose steps, use tools and take multiple actions to work towards a goal, with some ability to act on its own. It may involve one AI agent, several agents working together, or agentic capabilities built into another product.

An **AI agent** is software that uses agentic capabilities to carry out a task.

In this resource, we use agentic AI when talking about this type of AI system broadly, and AI agent when talking about an individual agent completing a task.

Understand what agentic AI can do

What is agentic AI?

Some AI systems answer questions or create content. Agentic AI can go further by planning, choosing steps, using tools and taking actions to complete a goal.

Rather than responding to a single prompt, agentic AI systems can gather information, decide what to do next and work through multiple steps towards a goal. Agentic features can be part of an existing product or service, or provided through a standalone AI system.

AI agents can have different levels of autonomy. Some may require a person to approve particular actions, while others can take actions without approval.

Example: How an AI agent might help

On the farm	At the salon	In the store
A farm might use an AI agent to monitor weather and soil data, adjust an irrigation plan when conditions change, and alert staff if something needs attention.	A hairdressing business might use an AI agent to manage bookings, find new times when an appointments change and update schedules.	A retailer might use an AI agent to monitor stock, identify when something is running low and then prepare a supplier order for approval.

How to recognise agentic AI

A simple way to identify agentic AI is to ask who takes the next step.

If the AI responds by providing information, content or suggestions, a person usually decides what happens next. This is not typically considered agentic AI.

If the AI can work towards a goal by using tools, making decisions and taking actions, it is likely to have agentic capabilities.

Agentic AI can take different forms

Agentic capabilities may be built into existing software, delivered through a single AI agent, or involve multiple AI agents working together.

Type	How it works	Example
Built into a product	Agentic capability may be built into software you already use, even if it's not labelled as an AI agent.	A booking system that automatically schedules appointments and sends reminders.
Single agent	An AI agent may carry out a task or work through several steps towards a goal.	An agent that manages customer enquiries and updates records.
Multiple agents	Several AI agents may work together, with different agents carrying out different parts of a task to achieve a goal.	Intake, drafting, quality assurance and compliance agents work together to create a newsletter.

Example: How multiple AI agents can work together

Some agentic systems use multiple AI agents to achieve a goal. Each agent is responsible for a different part of the work.

The example below shows how four AI agents could help prepare a newsletter.

- The **intake agent** receives the request, confirms its purpose and identifies the intended audience.
- The **drafting agent** uses this information to prepare a draft.
- The **quality assurance agent** checks readability, spelling, links and accuracy. If issues are identified, it requests changes.
- The **compliance agent** checks the draft against relevant brand, legal, regulatory and privacy requirements.

The agents share information and coordinate their work to prepare the draft for review. A person then reviews and approves the final version before it is published.

This example shows how multiple AI agents can complete different parts of a task while a person remains responsible for the final decision.

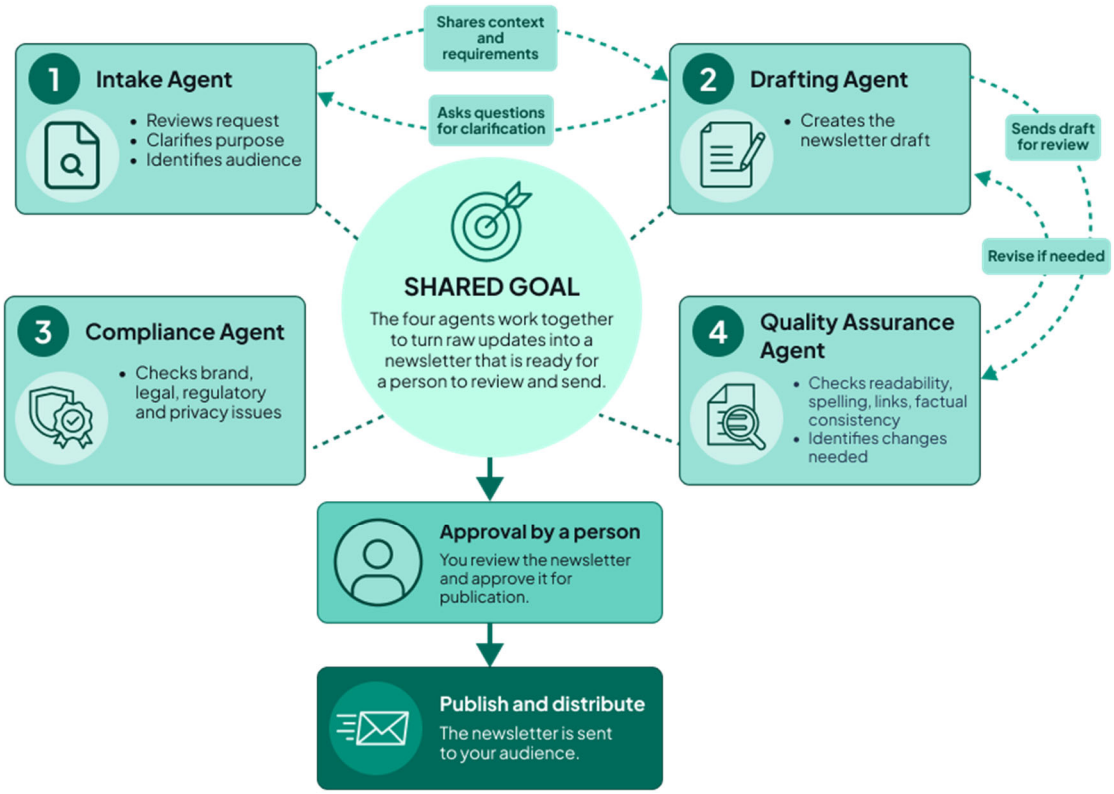


Figure 1: Example of how multiple AI agents can work together.

How can agentic AI help?

When agentic AI is used responsibly and effectively, it can help businesses save time, reduce manual administration and respond faster to customers, among other potential benefits.

For example, an AI agent might help a small retailer answer common customer questions, check stock levels, prepare invoices and flag anything unusual for a person to review.

Examples of how AI agents could be used in business

Business area	How the AI agent could help	Example
Customer service	Handle customer requests, check relevant information and take or prepare the next step.	A retailer uses an AI agent to respond to delivery questions, check an order's status and arrange an updated delivery when needed, while flagging complaints for a person to review.
Bookings and scheduling	Check calendars, suggest times, send reminders or update bookings.	A hair salon uses an AI agent to manage appointments and send customer reminders.
Stock and ordering	Check stock levels, suggest reorders or prepare supplier messages.	A café uses an AI agent to track low stock and prepare a supplier order for approval.
Monitoring and alerts	Watch for changes, unusual activity or missed steps.	A farm uses an AI agent to check weather and soil data and alert staff when irrigation may be needed.

Is agentic AI the right tool?

Before deciding to use agentic AI in your organisation, identify the problem you want to solve.

Ask:

- What problem are we trying to solve?
- What task or process needs improvement?
- Could a simpler tool achieve the same outcome?
- What are the costs and expected benefits?
- What could happen if the AI makes a mistake?
- Does the task involve customers, money, personal information or important business records?

If a simpler tool can achieve the outcome, you may not need agentic AI.

What job will it help with?

Write down the task you want agentic AI to help with in plain language and identify the main steps involved.

For each step, ask:

- Who could be affected if something goes wrong?
- Should the AI act on this step or provide recommendations only?
- Should a person approve this step before the next action is taken?
- Should this step always remain under human control?

- Which actions, if any, could the AI take on its own?
- Who would be responsible if the AI makes an error?

Thinking through each step can help you decide where human oversight is needed and where automation may be appropriate.

Who is responsible if something goes wrong?

Before using agentic AI, decide who is responsible for monitoring its actions, reviewing outcomes and responding if something goes wrong.

Consider the potential consequences of errors or unintended actions, including any legal, financial or reputational impacts.

This includes understanding how responsibilities are shared between your organisation and any technology providers.

Existing laws apply when you use AI and organisations using and developing AI remain responsible for identifying and responding to AI harms. Seek professional advice if you are unsure about your legal obligations.

Understand agentic AI risk

Like all technologies, organisations need to balance the potential value of agentic AI against the risks.

Many of the risks of agentic AI use are not new types of risk. But they can become more serious when an AI system can access sensitive information or act repeatedly, by itself, with other agents or across connected tools. Organisations still need to make sure they meet all regulatory obligations, regardless of whether a task is performed by a team member or agentic AI.

Common risks of agentic AI

The examples below identify some common risks associated with agentic AI and practical steps that may help reduce them. They are not a complete list of risks or risk controls.

Risk	What could happen	Steps to help reduce the risk
Costs can grow quickly	AI agents can be more costly than expected if they perform many steps, use multiple tools or continue working in the background. Costs can also grow quickly when agents are overused, given unclear instructions or operate without clear limits.	• Set spending limits. • Monitor usage and regularly review whether the agent is delivering value.
Problems may go unnoticed	Mistakes or changes in how an AI agent behaves may go unnoticed without regular review.	• Make sure someone is responsible for reviewing the agent's work. • Make sure they have the time and skills to regularly fix problems and make improvements.
Personal information may be exposed	AI agents with access to personal information, multiple systems and outside sources may increase the risk of unauthorised use or sharing of information.	• Only give the agent the access it needs for its approved task. • Review and remove access when it is no longer needed.
Errors can spread across tasks and systems	AI agents can make mistakes, behave unexpectedly or take actions that were not intended. Because AI agents can continue working across multiple steps, a single mistake can have a wide impact.	• Limit what the agent can do. • Test it before it can take actions on its own. • Regularly review its actions and results. • Make sure a person can stop or correct mistakes.
Important data may be changed or deleted	AI agents that can update files or records may accidentally change, overwrite or delete important information.	• Keep secure backups or copies of important information. • Test that you can restore the information if needed.

What information will it use and store?

Storing information can be useful where an agent needs it for future tasks. But you should carefully consider what should be kept, for how long, and whether the information remains appropriate and accurate over time.

Ask:

- What information does the agent need, including customer, staff or financial information?
- Where will that information be processed and stored?
- What systems or apps does the agent need to use?
- What information, files, instructions or settings will the agent keep?
- Can the agent's memory, instructions or settings change over time?
- How long will the agent keep information, and can humans completely delete it?
- How does the agent correct or remove information that is wrong or out of date?

Remember: An AI agent may identify or infer sensitive information by combining information from different places, even when each piece of information seems safe on its own.

What actions can agentic AI take?

With agentic AI, the risk is not just what the AI agent says. It is what it can see, change, send or do.

To get a more complete picture of the level of risk your agent could present, ask:

- Can it access business systems or files?
- Can it use tools, APIs or connected apps?
- Can it send messages, update records, place orders or trigger workflows?
- Can it make decisions or take actions without approval?
- Does it follow fixed steps, or can it decide what to do next based on changing factors?
- Can it interact with customers, suppliers or external systems?
- Can it connect with other AI agents?
- Can it act outside its approved limits, change those limits or gain greater access?

The more questions you answer 'yes' to, the stronger your controls should be.

Remember: Just because information or access is technically available does not mean the AI agent is authorised to use it. External websites, systems and AI agents may also provide unreliable information or influence how an agent behaves.

What access should an AI agent have?

Set clear limits on what an AI agent can access and do.

Check what information the agent can access through connected tools and systems, not only the information you provide directly.

A good rule is to treat an AI agent like a new staff member or contractor. If you would not provide them with access to information, systems or tools without appropriate checks, do not give that access to an AI agent.

Examples: matching access to the task

On the farm	In the salon
An AI agent that helps with irrigation may need access to: • Weather data • Crop information • Soil moisture data	An AI agent that helps with bookings may need access to: • Appointment information • Staff availability • Customer booking requests
It may not need access to: • Invoices • Private staff records • Banking information	It may not need access to: • Payroll information • Customer payment details • Private staff records

What can increase agentic AI risks?

Some factors can increase the likelihood or impact of mistakes, security issues or unexpected behaviour.

One mistake can lead to another. AI agents can repeat the same mistake, take unexpected shortcuts or pursue goals in unintended or harmful ways. It may also be difficult to understand exactly how a mistake happened.

When multiple AI agents interact

Agentic systems sometimes use multiple AI agents to complete different parts of a task.

While each AI agent may work as intended on its own, different risks can emerge when agents interact, share information, pass work between each other or interact through shared systems and tools. A mistake made by one agent may be repeated, amplified or passed on to other agents.

When multiple AI agents can access the same systems, tools or information, they may influence each other's actions in unexpected ways, even if they were not designed to work together.

The more agents interact, the more opportunity there is for unexpected behaviour.

Check how the agents, tools and connected systems work together, not just whether each component works as intended on its own.

For more information, see the Australian AI Safety Institute's Multi-Agent Risk Report.

Three factors that can increase risk

Risks can also increase when an AI agent can combine information from different sources and use it to take actions on behalf of a person or organisation. If the information is incorrect, manipulated or sensitive, the impact of a mistake may be greater.

An AI agent may present higher risks when it can do all 3 of these things together:

1. Access personal, sensitive or business information
2. Obtain information from external sources
3. Take actions on behalf of a person or organisation.

Cyber security considerations

If the AI agent will connect to sensitive information, critical business systems or external services, consider seeking cyber security advice before using it.

The Australian Cyber Security Centre (ACSC) guidance on the careful adoption of agentic AI services provides more detailed cyber security considerations.

The Australian Cyber Security Centre (ACSC) guidance on agentic AI harnesses also includes practical examples of controls that can help limit what an AI agent can access, what it can do and when human intervention is required.

Use agentic AI safely

When using AI, including agentic AI, it is good practice to have responsible systems and processes in place. Organisations also need to ensure they have processes in place to meet existing regulatory obligations.

The Guidance for AI Adoption sets out 6 practical steps to help organisations build good foundations for responsible AI adoption and manage risks. Organisations using agentic AI should check that they have these 6 key practices in place.

Work health and safety laws apply when using AI and AI agents, as they do with other tools used in the workplace. Safe Work Australia provides information on WHS duties and managing risks associated with AI and digital technologies.

Use agentic AI on a small scale first

Start with one small, clear task for agentic AI and test it. Choose a task that is useful enough to show whether the AI agent helps, but limited enough that problems are easy to spot, contain and fix.

Starting small does not remove risk. It limits the consequences and makes problems easier to detect while you test. You still need to use appropriate access limits and monitoring during development and testing.

Examples of how to start small and safely test your use of agentic AI:

- Run a short pilot with restricted access
  Test the AI agent with a small number of staff, a clearly defined task and access only to the information needed for that task.
- Test in a low-risk setting first
  Use test data, sample records or a copy of a process before connecting the AI agent to live business systems.
- Start with draft-only or read-only access
  Allow the AI agent to prepare work, suggest next steps or identify issues before it is permitted to take actions like changing records, sending messages, placing orders or triggering workflows.
- Set clear stop points
  Decide in advance when a person must review the work, approve the next step or stop the test.
- Check how it handles challenges
  Test how the AI agent responds to mistakes, unusual requests and attempts to act outside the approved task.
- Review results before expanding use
  Check whether the AI agent completed the task as intended, whether staff understood what it did, whether problems were easy to identify and fix, and whether the benefits justify broader use.

The goal is to control the risks of agentic AI and assess whether it delivers value for the task.

When to pause and get advice

If the AI agent will access sensitive information, handle payments, connect to critical business systems or take actions that might be hard to reverse, consider seeking cyber security, privacy, legal or technical advice before proceeding.

Pause and get advice from the relevant expert or advisor if the AI system can access or act on:

- financial information, including bank accounts and payment systems
- passwords
- staff or customer records
- health, or legal information
- private files
- emails
- publishing public-facing content such as on websites or social media channels
- orders, invoices or refunds
- systems where mistakes are hard to undo
- core systems or data, including the ability to delete these.

Providing agentic AI with access to these systems and types of information may still be possible. But before doing so you will need stronger checks, clear responsibility and a way to stop or reverse actions.

The more systems, organisations and AI agents involved, the harder it becomes to manage risk. Stronger checks may be needed when AI agents interact with:

- customers
- suppliers
- business partners
- external systems
- other AI agents.

The more of these capabilities an AI agent has, the more important it is to set limits, keep people involved, monitor what it does and make it easy to stop.

Four safety checks before you turn it on

No single safeguard is enough. Safe use of agentic AI depends on a combination of people, processes and technical controls.

Before you use an AI agent, pay close attention to these 4 things:

- Access
  Give the AI agent access only to the information, systems and tools it needs to complete its task. Limiting access reduces the potential impact of mistakes and helps ensure the AI agent can only perform approved activities.

- People
  Keep people involved and accountable where it matters. Decide when a person should review, approve or stop an action, particularly where decisions could affect customers, staff, finances or important business processes.

- Visibility
  Make sure you can see what the AI agent is doing. You should be able to understand what information it used, what actions it took and what changes it made. This can help you identify problems, investigate incidents and improve performance over time.

- Stop
  Make it easy to stop an AI agent if needed. Before deployment, understand how to pause the AI agent, remove access, stop actions and recover from mistakes where possible.

Why human oversight is not always enough

Having a person review an AI agent's actions can help reduce risk, but it is not always enough on its own.

An AI agent may operate at a speed or scale that makes it difficult for a person to review every action. Human oversight is most effective when the reviewer has enough time, information, expertise and authority to intervene when needed.

This is why the four checks work together:

- Access limits what the AI agent can see and do.
- People helps ensure important actions receive appropriate review.
- Visibility such as records of significant actions helps you understand what the AI agent did
- Stop makes it possible to pause the AI agent, remove access or intervene if something goes wrong.

As AI agents become more capable or operate with greater autonomy, relying on human review alone may become less effective. Additional safeguards, such as systems that automatically deactivate agents that breach controls, may be needed to help manage risk.

Manage unapproved use of agentic AI

Staff may already be using personal AI tools for work. New agentic capabilities may also be added to software your organisation already uses, making them easy to overlook.

When AI tools are used without organisational visibility or approval, it is sometimes referred to as shadow AI.

Keep track of which AI agents and agentic features are being used across your organisation. Understand what they can access, what actions they can take and who is responsible for them.

Rather than relying only on restrictions, consider providing safe, approved tools that meet business needs. This can help reduce unapproved use while giving staff access to tools that have been assessed and supported.

Learn more about managing unapproved AI use.

Ask your vendor the right questions

Many organisations will use agentic AI through products and services provided by vendors.

Different products can work in different ways, and it is not always easy to understand how an AI agent will behave in practice from product information alone. Ask questions and request clear answers before purchasing, enabling or expanding the use of an agentic AI product.

Consider asking:

- What information will the AI use?
- Where is our data stored?
- Will our data be used to train the AI?
- What systems can it access?
- What actions can it take?
- Can we limit what it can do?
- Can we see logs of what it did?
- Can we pause or turn it off?
- What happens if it makes a mistake?
- Can access be limited or automatically removed when no longer needed?
- Can costs rise through automated use?
- What information, files or settings will it keep?
- Can we see what it accessed, changed, sent or decided?
- Can we tell what changed and when?
- How do we know if the product or model changes?
- If it changes or deletes information, how do we restore it?
- How do we know it has actually stopped?
- What does the agent do if a connected tool or system fails?
- Has the agent been designed to comply with all relevant legal requirements?

- Who is responsible for what if something goes wrong or the AI causes harm, including when it connects with another provider, system or AI agent?

The answers to these questions can help you identify risks, put appropriate safeguards in place and determine whether any existing policies, procedures, approvals or governance arrangements need to be updated before deployment.

Keep your AI agent under review

Using an AI agent is not a one-time decision. As the AI agent, the task, the information it uses or the systems it connects to change, the risks may change too.

Review the AI agent when significant changes occur. Keep a record of what the agent can access, what information it keeps, who approved the change and when it was made. This can help you understand how the AI agent has evolved over time and support investigations if something goes wrong.

Examples of changes that could trigger a review include:

- changes to the AI model, product or vendor
- changes to the information it uses
- changes to the task it performs
- access to new systems, tools or data
- additional staff using the AI agent
- customers being affected by its actions
- connections to new systems, services or other AI agents
- a complaint, error or near miss
- behaviour outside its approved purpose or role
- identification of a workplace health and safety hazard.

Make sure someone in your organisation is responsible for the performance of each agentic AI system in use. They need to monitor how it performs, detect problems, limit the impact and respond when something goes wrong.

Before you use agentic AI

Using agentic AI is not a one-time decision. Before you deploy an agentic system, take time to understand the task, identify potential risks, put appropriate safeguards in place and decide who will be responsible for its use.

The checklist below can help you assess whether an AI agent is appropriate for the task, what safeguards may be needed and whether any existing policies, procedures, approvals or governance arrangements should be updated before deployment.

Agentic AI checklist

1. Define the task

Before selecting or configuring an AI agent, be clear about what you want it to do.

☐ What tasks will the AI agent help with?
☐ Is an AI agent the right tool for those tasks?
☐ Who could be affected if something goes wrong?
☐ Have we identified the benefits we expect to deliver?

2. Understand the risks

Understand what the AI agent can access, what it can do and the potential impact of mistakes.

☐ What information will it use or keep?
☐ What systems, tools or data can it access?
☐ What actions can it take?
☐ What actions should require person approval?
☐ Have we considered who may be responsible or liable if something goes wrong?
☐ Have we considered any work health and safety impacts?
☐ Are appropriate controls in place to manage the risks?

3. Put controls in place

Make sure the AI agent operates within clear limits.

☐ Does it have only the access needed for the approved task?
☐ Can we see what it accessed, changed, sent or decided?
☐ Can we stop it quickly if needed?
☐ Can we detect problems, respond to them and recover from mistakes where possible?
☐ Who is responsible for monitoring the AI agent and acting if something goes wrong?

4. Prepare for deployment

Before expanding use, confirm that the AI agent is suitable for your organisation.

☐ Have we started small and tested it?
☐ Have we asked vendors the questions we need answered?

☐ Do we know which AI agents and agentic features are active?
☐ Do our policies, procedures, approvals and governance arrangements support this use?

5. Keep it under review

Review the AI agent regularly, including when its use, access or impact changes.

☐ Do we know when the next review will occur?
☐ Do we have a process for reviewing significant changes?
☐ Do we know who is responsible for ongoing oversight?
☐ Do we know how issues, incidents and near misses will be reported and managed?