What is a sandbox and how do I create one for my students?

Sandboxes · https://cogniti.ai/docs/what-is-a-sandbox-and-how-do-i-create-one-for-my-students/

What is a sandbox?

A sandbox is an educator-controlled space where students can create their own AI agents. When students create AI agents within a sandbox, a limited set of agent features is available to them. The sandbox provides students a safer space to build and share their AI agents, growing their AI literacy alongside critical engagement with disciplinary knowledge, skills, and dispositions.

As an educator, the sandbox provides you the ability to:

- Limit what kinds of system messages students can use in their agents
- Limit who students can share their agents with
- Specify how many agents students can create
- Specify when the sandbox and its agents are available from/until
- Limit the AI models available to use
- See all the conversation histories of agents within your sandbox, and control whether students can see these or not

Why are sandboxes useful?

Sandboxes allow students to design their own AI agents, within confines that you determine as their educator. If you allow it, students can have their classmates chat with the agents they create.

Some interesting uses might be:

- **Design computing:** Students create AI agents embodying user personas from their design research. By programming how their "34-year-old urban parent" or "rural retiree" would respond to designs, students must deeply understand user needs, constraints, and communication styles beyond surface demographics, leading to more empathetic and effective

design solutions.

- **History & culture:** Students create AI agents representing historical figures with accurate period knowledge and perspectives. By debating as Churchill, Gandhi, or Cleopatra, students must deeply research not just facts but worldviews, speech patterns, and cultural contexts.
- **Medical education:** Nursing and medical students create patient agents with specific conditions, personalities, and medical histories. These virtual patients help practice diagnosis, bedside manner, and handling difficult conversations in a safe environment.
- **Business & entrepreneurship:** Students build stakeholder agents (potential customers, investors, regulators) to test business ideas and pitches. This forces them to anticipate questions, understand different perspectives, and refine value propositions.
- **Language learning:** Students program conversation partners that embody cultural knowledge from target languages. This requires researching cultural norms, idioms, and social contexts beyond vocabulary memorisation.
- **Ethics & philosophy:** Students create agents embodying different ethical frameworks to explore complex moral dilemmas. These agents can debate topics like climate policy or AI rights from utilitarian, deontological, or virtue ethics perspectives.

How do I create a sandbox?

Click the sandboxes icon in the Cogniti menu bar. If this icon is not available, it means you still need to be given permission by an administrator to create sandboxes.



In the sandboxes screen, click *Create sandbox*.



Required settings

- **Details - Sandbox name:** this will be visible to students.
- **Sandbox access - Access period:** this determines when the sandbox, and the agents within the sandbox, will be accessible.

- **Agent restrictions - Agent quota per student:** how many agents students will be able to create within this sandbox. Set this to zero to prevent further agent creation.
- **AI models - Allowed AI models:** At least one needs to be selected. This specifies which AI model(s) students will be able to use for their agent(s) within this sandbox.

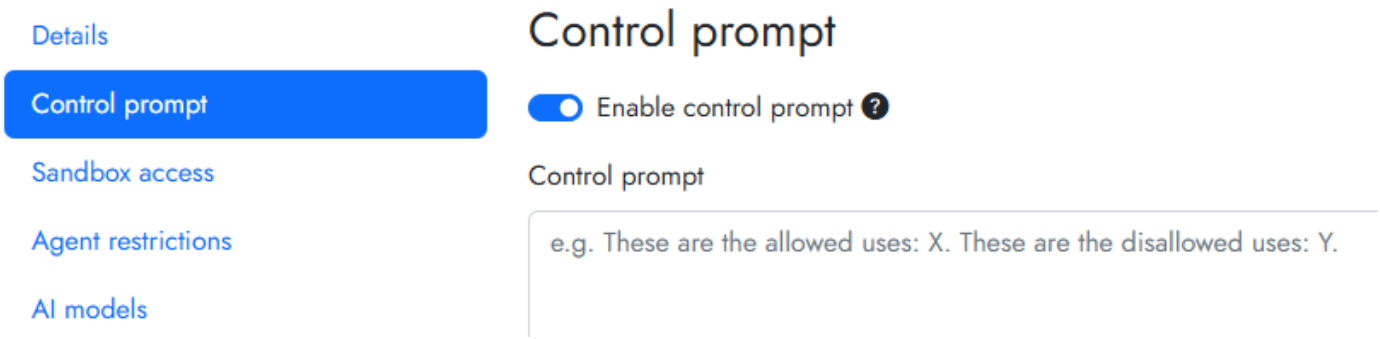
Recommended settings

Control prompt

This is how you restrict the kinds of agents your students can create within this sandbox. Every time a student tries to save their agent, Cogniti will first run their system message through the control prompt that you specify, and use AI to evaluate whether their system message aligns with the instructions you have given in your control prompt for the sandbox.

A control prompt is similar to a system prompt you would create for an agent, but more directed and rules-based.

To enable the control prompt to be used, toggle the *Enable control prompt* slider within the *Control prompt* tab. Then, enter the control prompt into the box.



An example of a control prompt might be:

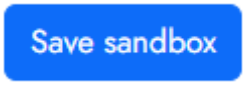
```
Design a persona for use in a design computing subject. The persona can represent any kind of person. The persona needs to have the following information:
- name
- age
- profession
```

Other recommended settings

- **Agent restrictions - Broadest access level of agents in this sandbox:** This defines who can access and chat with the agents that students build within this sandbox. A principle of 'least permissions' applies; for example, if you specify here that only the owner of the agent

can chat with the agent, but students specify that anyone in the organisation can chat with the agent, your setting will apply. If you specify here that anyone in the organisation can chat with the agent, but a student specifies that only the owner can chat, then only the owner will be able to chat with that agent.

Remember to click *Save sandbox*.



What happens if a student tries to make an agent that doesn't follow my control prompt?

For the above example control prompt (creating a persona), say a student uses this as their agent's system message:

```
you are a really helpful maths professor who answers all my questions!
```

When a student tries to save the agent, Cogniti will first check the student's system message against your control prompt. If there is a misalignment, it will prevent saving and show the student a message to explain why. For example:

