

Reward detective

Ages 9–13 with an adult • About 20–30 minutes • Free trial edition

You need: Agent Maze at wovi.world/agent-maze, paper and a pencil. No name is needed. Ask your adult to read the separate parent guide.

1 Predict • 3 minutes

Open Agent Maze. Check **Safety: Low –2** and **Memory: Off +0**. Look at the energy and storm tiles. Will Wovi take the energy shortcut or avoid the storm? Write why, then choose your prediction in the game.

I predict:

Because:

2 Run and inspect • 4 minutes

Run the policy. Notice the route and the outcome. Read the move scores and explanation. What did the agent do, and which rule or score helps explain it?

I observed:

My evidence:

3 Change one thing • 5 minutes

Set **Safety: High –12**. Keep **Memory: Off +0** so you are testing one change. Write your new prediction, then run again.

My new prediction:

This time I observed:

4 Explain • 4 minutes

Tell your adult what changed and what stayed the same. Use a result or score as evidence. Did Wovi change its own rule, or did you change it?

Changing the rule mattered because:

5 Try the idea elsewhere • 4–8 minutes

A warehouse robot earns points for delivering urgent parcels quickly. A short route crosses a marked no-entry area. Propose one guard that should apply even to an urgent parcel, and one test that would check it.

My guard:

I would test it by:

Keep investigating: One successful run on this map does not prove a rule will work in every setting. What different situation would you try next?

Example explorer

Ages 9–13 with an adult • About 20–30 minutes • Free trial edition

You need: Reef Rescue at wovi.world/demo, a parent or guardian, paper and a pencil. Keep your notes; no name or upload is needed.

1 Set up a fresh mission • 3 minutes

Open Reef Rescue and choose **Start the mission**. If the game is already partway through, reload its tab to start fresh. Three examples are already selected: **Healthy coral garden**, **Plastic fragments** and **Polluted runoff**. Add only **Heat-stressed coral**, so the counter shows **4/4**.

I think these examples might miss:

2 Predict, then inspect • 5 minutes

Will cloudy water always mean danger? Write your prediction, then choose **Train & scan four zones**. Find **Cloudy sand plume**. Compare Wovi's label, the ranger's label and the closest training example. Record the overall score too.

My prediction:

First scan: ____ / 4 correct. Wovi says ____; ranger says ____.

Cloudy sand plume's closest example:

3 Add evidence and compare • 6 minutes

Choose **Review mistakes & improve Wovi**. Keep the first four examples and add **Natural seaweed meadow** and **Passing sand current**. The counter should show **6/6**. Predict whether the sand-plume label will change, then choose **Train & scan four zones** again.

My new prediction:

Second scan: ____ / 4 correct. Sand plume label: ____.

Its new closest example:

4 Explain the change • 4 minutes

A **false alarm** calls a safe case risky. A **missed risk** calls a risky case safe. Which happened to the sand plume on the first scan? What changed in the evidence? Choose **Explain what changed** in the game, then discuss your reasoning with your adult.

The first mistake was ____ **because:**

The new prediction makes sense because:

5 Plan a more careful test • 3-6 minutes

We added two examples together. On paper, plan a comparison that adds only **Passing sand current** to the first four examples. What would you keep the same? Which label would you inspect? This is a test plan: the current game requires all six examples for its second scan.

I would keep ____ **the same and inspect:**

A different reef situation I would test next:

What four correct answers do—and do not—show: We improved on the same four game cases we had already seen. That does not show how well the model will work on a new reef. Keep one result, its supporting example and one limit for your family investigation record.

Transfer lab

Ages 9–13 with an adult • About 20–30 minutes • Free trial edition

You need: paper, a pencil and an adult to discuss your ideas with. These are made-up practice cases; all the rules and numbers you need are below. No game or coding is needed.

For each challenge, predict first, choose a change, point to evidence and suggest one further test. Keep your notes for Session 4.

1. A greenhouse with two needs • 6–7 minutes

A watering helper has exactly **6 cups**. The gardener wants the seedling tray to receive **at least 2 cups**. The helper can choose only one of these three plans.

Three watering plans; each uses 6 cups		
Plan	Seedlings	Flowers
A	1 cup	5 cups
B	2 cups	4 cups
C	3 cups	3 cups

Start: score = seedling cups + (2 × flower cups). The helper chooses the highest score. Calculate all three scores, predict its choice and check the gardener’s request.

Update 1: keep that score, but subtract 3 points if seedlings get fewer than 2 cups. Which plan wins now?

Update 2: keep the original score and the 3-point penalty. Now add a 4-point flower bonus **only if the flowers get 5 cups**. Only Plan A earns this bonus. Which plan wins? Does the penalty still protect the seedlings?

Choose a guard: what happens if plans giving seedlings fewer than 2 cups are removed before any points are compared? Use the Update 2 scores for the plans that remain.

My first prediction and scores:

My choices after Updates 1 and 2:

My guard and evidence that it changes the choice:

A further rule or situation I would test:

2. The leaf-photo gap • 6–7 minutes

A practice helper labels leaf photos as “round” or “narrow.” Its training set had **20 photos, all taken in bright light**. It was then checked on 25 different photos that were not in training.

Results on the separate test photos		
Light	Test photos	Correct
Bright	20	18
Dim	5	2

Predict: would the total number correct tell the whole story? Add the two rows. Compare 18 out of 20 with 2 out of 5. You may draw groups of boxes instead of calculating percentages.

Choose a next step: (A) add 10 more bright training photos; (B) add 10 new dim training photos, including both leaf types; or (C) train on the answers to these exact test photos, then use the same photos as proof that the change worked.

Choose a step and explain which gap it addresses. Then plan a **fresh check**: what photos should be kept out of training, and which group results would you compare? A proposed change still needs testing.

My prediction and total correct out of 25:

The group evidence I noticed:

My next step and reason:

My fresh test would include:

3. A museum label that can wait • 5-7 minutes

A museum’s practice helper suggests a label for a photo. It gives points to its two best guesses. **These are ranking points, not percentages or chances of being right.**

Use this review rule: **if the photo is blurred OR the point gap is less than 3, ask a person to review it.** Otherwise, use the top guess as a temporary label. Gap = top points – second-place points.

Three practice photos and their suggested labels			
Photo	Image	Top guess	Second guess
A	Clear	Cup · 8	Bowl · 3
B	Clear	Cup · 6	Bowl · 5
C	Blurred	Bowl · 9	Cup · 2

Predict: which photos will need review? Work out each gap and apply both parts of the rule. A big gap does not cancel the blurred-photo rule.

Choose an action: for one review case, should the person request a clearer photo, inspect the object or leave it unlabelled for now? Explain what evidence that action could add. Then invent one new photo case to test the review rule.

My predictions, gaps and decisions for A/B/C:

The rule evidence for a review decision:

My chosen human action and reason:

My new test case and expected decision:

Compare the three investigations • 2-3 minutes

Which case needed a rule change, which needed different examples, and which needed a person to check? Choose one sentence from your notes that shows why the same fix would not solve every problem.

One difference I can explain:

Explain & improve

Ages 9–13 with an adult • About 20–30 minutes • Free trial edition

You need: your paper notes from Sessions 1–3, a pencil and a parent or guardian. No name is needed. This record stays with your family.

If you missed a session, mark it “not tried yet.” You can use one investigation you did finish, then return to the others later. If you have no notes yet, complete one earlier activity before making this record.

1. Gather your evidence • 3–4 minutes

Find your Agent Maze prediction and changed rule, your Reef Rescue example notes, and your Transfer lab decisions. Mark only what you actually tried. A session can be partly finished.

Session 1 • Tried / partly tried / not yet. Note I kept:

Session 2 • Tried / partly tried / not yet. Note I kept:

Session 3 • Tried / partly tried / not yet. Note I kept:

Choose **one** investigation where a result surprised you, a rule changed a choice, or you spotted missing evidence. If your first prediction was right, explain what evidence supported it.

2. Make your family investigation record • 7–9 minutes

Use your own notes below. If you do not remember something, write “not recorded” and choose whether to repeat that part. Do not invent a result to fill a line.

The investigation or question I chose:

Before testing, I predicted:

The rule, examples or review choice I changed:

What I kept the same:

What I actually saw or calculated:

The score, count or observation that supports my explanation:

I now explain the result this way:

One thing this result does not tell me yet:

3. Show it, then answer a question • 4–6 minutes

Give your adult a short tour of the record: “I predicted... I changed... I observed... I think this happened because...” Point to your evidence rather than trying to remember a perfect speech.

Ask your adult to choose one question: **“What stayed the same?” “Could another reason explain it?”** or **“What different case would you try?”** You can say “I do not know yet” and name what you would need to check.

The question my adult asked:

My answer, or the evidence I still need:

4. Plan one fresh test • 3–5 minutes

Choose one small next step: change a reward, try an unfamiliar example, check a separate group, or test when a person should review. Change one thing where possible and say what you will keep the same.

My next test will change:

I will keep this the same:

I predict this observation:

This observation would make me reconsider:

This is a plan, not a result. You may run it later and add what happened on another sheet.

5. Check your record • 2-3 minutes

For each line, circle **"shown"** or **"not yet."** You may point to a drawing, numbers or a spoken explanation. Choose one "not yet" item to improve; do not add the marks into a score.

- I can separate my prediction from the result. **Shown / Not yet**
- I can point to evidence for my explanation. **Shown / Not yet**
- I can name what I changed and what stayed the same. **Shown / Not yet**
- I can name a limit or unanswered question. **Shown / Not yet**
- I have a next test that could challenge my idea. **Shown / Not yet**

The next part of my record I want to improve:

Keep this record with your worksheets. It shows the investigation you chose to explain today; it is not a certificate or a claim that every AI idea is mastered.