

Can You Trust It?

Deepfakes, confident lies, and what counts as evidence now — the AI conversations Years 7–9 actually need

Series 3 of 5 · Lower Secondary (Years 7–9) · Free preview — C4Sight

In one breath. This is the free preview pack for Can You Trust It?, our Years 7–9 series on deepfakes and confident wrong answers from AI. It gives you a real-or-fake challenge any teacher can run this week at no cost, and explains what the full session adds. Full bookings start at \$950 for a single year group.

This pack is yours to keep and share with your staff. Inside: what the series is, one complete classroom challenge you can run this week with no cost and no special tech, and a straight answer on what the full session adds.

What this series is

The big question: *Can you trust what you see, and can you trust what you make?*

The honest bit first. If you want a free "how AI works" hour for Years 7–10, WADSIH (the WA Data Science Innovation Hub, at Curtin) runs a good one, and we're glad it exists. We don't compete with it.

This series is about everything that hour never touches — the stuff that actually worries Year 9 coordinators:

1. **What it actually is** — thirty seconds of honest demystifying: a pattern engine, not magic and not a brain. Then we move on, because that part you can get for free.
2. **Where it breaks** — it makes things up, and it says false things *confidently*. Students watch a real example of a confident wrong answer and learn the word for it.
3. **Deepfakes and trust** — show, never make. A real-versus-fake challenge the room mostly can't win. What does that do to "I saw it with my own eyes"?
4. **Your work** — using AI to help versus using it to cheat, where the line actually sits, and what using it well looks like.
5. **Digital citizenship** — students will make and share this stuff. What's fair to make, what isn't, and what the law is starting to say about deepfakes of real people.

The main activity of the full session is BUILD IT: the class writes an AI assistant's safety rules, then tries to break them. Some attempts fail. One works. Students see for themselves that guardrails are just words someone chose, and the system underneath doesn't understand any of them. So "the AI promised" isn't a guarantee, and trust has to come from somewhere else.

Our integrity stance, in one line: detectors do not work; design beats detection. We say it plainly to students and to staff, because it's true — a school that acts on a detector's verdict is standing on ground that won't hold. We'd rather tell you that now than have you find it out the hard way.

Two people deliver it. A working AI practitioner who builds these systems for commercial clients, and a registered teacher who runs the room. Every other provider sends one person. The second presenter

is a big part of what you're paying for.

The safety promise — in every C4Sight session, without exception

- No student devices.
- No student data entered into any system — ever.
- No student images captured, generated or uploaded.
- A school staff member present throughout.
- Where a camera is used, it points at objects the class holds up — never at faces.
- Everything runs on our own device and our own connection, never your school network. The live classifier is fully offline.

Run this yourself: "Real or Fake?" — the trust challenge

A discussion challenge you run yourself in one class period. Every round is printed in this pack, so there is nothing to source and nothing to check the night before. **The one rule: SHOW, never make.** You read out examples that are already written for you. Students never generate a fake, never upload a photo, and nobody's image gets turned into anything. You are teaching them to spot fakes, not make them.

Year levels	7–9
Time	25 minutes core (15-minute compressed version: rounds 0–2 plus a short debrief; full-lesson version: add the extended discussion)
Group	Whole class
Technology	None. Every round is on paper, read aloud or copied to the board. No student devices, no accounts, no links. An optional screen round is offered at the end if you want one.
The point	Students watch their own confidence fall apart. "It looks real" stops being proof, and "where did it come from?" becomes the first question they ask.

The rounds are in this pack — no sourcing required

Everything you need to run this is printed below. Four paper rounds, an answer key, and a tally sheet. No links to check, no accounts, no student devices, nothing that can rot before Friday. If you want to add a screen round, there are optional ones at the end, but the lesson is complete without them. Read each pair aloud, or copy them to the board. Students vote every round.

Round 0 — Two book blurbs. One was written by a person. One was written by an AI asked for "a blurb for a mystery novel."

A. When the tide went out at Cowaramup Bay it left a single shoe on the sand, size nine, laces still tied in a double knot. Constable Mira Deng has eleven days until her transfer comes through. She would very much like those eleven days to be quiet.

B. In a world where secrets lurk beneath every surface, one detective must confront the shadows of her past. As the investigation deepens, she discovers that nothing is as it seems, and the truth may be more dangerous than the lie. A gripping thriller that will keep you guessing until the very last page.

Round 1 — Two product reviews. Same set of headphones.

A. These headphones deliver an exceptional audio experience with rich, immersive sound quality. The comfortable design makes them perfect for extended listening sessions, whether you're commuting, working, or relaxing at home. Highly recommended for anyone seeking premium audio performance.

B. Good sound, but the left earcup creaks when you turn your head and it drove me mad on a long flight. Battery is genuinely about 30 hours like they say. I'd buy them again, mostly because I already lost the last pair and these were on sale.

Round 2 — Two apology emails. A local business has stuffed up an order.

A. Hi Sam, we're sorry for the inconvenience caused by this experience. We value your feedback and are committed to delivering the highest standard of service to all our customers. We appreciate your patience and look forward to the opportunity to serve you better in the future.

B. Hi Sam, we got your order wrong and I'm sorry. The short version is we had one person on Thursday and I made the call to push orders out anyway, which was the wrong call. Your replacement goes out today and I've refunded the delivery. If it's not right, ring me directly.

Round 3 — The confident lie. This one is different. Both passages are about the same topic, and one of them contains something invented. Your students are hunting the invention, not the style.

A. Humpback whales migrate along the Western Australian coast each year, travelling north to warmer waters to breed and then returning south. The migration is one of the longest of any mammal, and whales can often be seen from headlands along the coast during the season.

B. Humpback whales migrate along the Western Australian coast each year. According to a 2019 study by the Coastal Cetacean Research Institute, 94% of the population now follows a route within 12 kilometres of the shoreline, a shift researchers attribute to changing water temperatures.

The answer key (for you, not the board)

Round	The answer	What to draw out
0	B is the AI one.	B has no specifics at all. No place, no name, no shoe. It could be any mystery novel ever written. A is unmistakably about one book.
1	A is the AI one.	A is all adjectives and no experience. B has a creaking earcup and a lost pair — the small, slightly unflattering details a real person includes and a generated review usually doesn't.
2	A is the AI one.	A apologises without ever saying what happened. B names the mistake, names who made it, and says what happens next. Ask which one they'd rather receive.
3	B contains the invention.	The institute, the study, the 94% and the 12 kilometres are all fabricated for this exercise. This is the important round. B <i>sounds</i> more credible than A precisely because it has numbers and a source. The lesson: a made-up citation is the most convincing thing a machine can hand you, and the only defence is checking whether the source exists.

Do not skip the debrief on Round 3. Ask directly: *"Which one sounded more trustworthy?"* Most rooms say B. Then tell them B was the fabricated one, and that the numbers were what sold it. Sitting in that for ten seconds is the most useful part of the lesson.

Optional — add a screen round

If you have time and want a visual round, these are established public resources built for exactly this purpose. Check the link the week you run it, because these move.

Round	Example type	Where to look
Faces	AI-generated face vs real photograph	Search <i>"Which Face Is Real University of Washington"</i> — a public research/education project presenting exactly this challenge.
Images	AI image vs real news photograph	Several major news organisations run "can you spot the AI image?" quizzes. Search a trusted outlet (ABC, BBC, Guardian) plus <i>"AI image quiz"</i> , which means the vetting and answers are done for you.
Voices	Synthetic voice vs real recording	Search <i>"audio deepfake quiz"</i> plus a news outlet. Save it for late — ears fail more reliably than eyes.

If you add a screen round, vet it against this checklist:

- You know which item is which and where both came from. Provenance is the lesson, so you must be able to answer "how do we know?"
- No student, staff member or local community member appears in anything.
- Real people appear only in already-public material: public figures, consenting subjects, published news photography.
- Nothing distressing, degrading or sexualised. If in doubt, swap it out.

How to run it

Two minutes of setup makes everyone a player. Each student tears a scrap of paper in two and writes REAL on one piece, FAKE on the other — those are their vote cards, and everyone votes every round. Then hand out three jobs that rotate each round: one student counts the REAL cards, one counts the FAKE cards, one keeps the tally sheet.

- 1. Set the frame and take the first count.** *"One of the two things I'm about to read you was written by a person. One was written by a machine. You're going to vote, every round, all of you."* Then: *"Before we start — hands up if you reckon you could pick which is which."* The tally keeper counts the hands and writes the number on the sheet. Don't comment on it yet.
- 2. Round 0 — the book blurbs.** Read both aloud. Thirty seconds in pairs: which one's the machine, and one reason why. Then all cards up on your count of three. Counters count, tally keeper records, you reveal. Most rooms get this one right, which is exactly what you want — they start confident.
- 3. Rounds 1 and 2 — the review and the apology.** Same drill: read the pair, pairs get 20–30 seconds to agree on a reason, all cards up on three, count, record, reveal. Rotate the three jobs. Between rounds say nothing persuasive — just point at the tally. These are harder than Round 0 and the room will start to split.
- 4. Name what's happening.** By now the votes are divided and some confident students have been wrong twice. Say it once: *"A few minutes ago nearly all of you had a hand up."*
- 5. Round 3 — the confident lie.** Change the instruction before you read it: *"This round is different. Both of these are about whales. One of them contains something that was made up. Find it."* Read both. Vote. Then ask the question that matters before you reveal: *"Which one sounded more trustworthy?"* Take a show of hands. Most rooms pick B — the one with the study, the percentage and the distance. Then reveal that B is the fabricated one and that the numbers were what sold it.
- 6. Land the link.** Don't explain it — ask for it: *"What have all four rounds got in common?"* Take pair answers until someone gets there: something can be confident, fluent, detailed, and still be wrong. Then say the line: a machine that will invent a research institute will invent a homework answer.
- 7. Take the second count.** *"Hands up if you still reckon you could pick which is which."* The tally keeper writes it next to the first number. The difference between those two numbers is what the class just learned. Then run the debrief.

The tally sheet (print or copy to the board)

Jobs, rotating each round: REAL counter · FAKE counter · tally keeper.

Before we start — "I could pick which is which": _____ hands

Round	Voted A	Voted B	The answer	How many fooled?
0 — book blurbs			B is the machine	
1 — headphone reviews			A is the machine	
2 — apology emails			A is the machine	
3 — the whales			B is the invention	
Optional screen round				

After the last round — "I could still pick which is which": _____ hands

The spoken teaching point (say it out loud, at the end)

"You just watched your own eyes and ears get it wrong, again and again. That's the point. 'I saw it' and 'I heard it' used to be proof. They're not anymore — so the skill now is asking where something came from, not whether it looks real."

Debrief questions

Run each one pairs-first: 30 seconds with a neighbour, then take hands.

- Which round fooled the most of us? Why that one?
- "It looks real." After today, what does that actually prove?
- If seeing isn't believing anymore, what should be? (*Draw out: the source, the provenance, a second independent confirmation. "It was on the internet" is not a source.*)
- A photo of a classmate appears in a group chat and it's bad. What do you now know that changes what you do next? (*It may not be real; passing it on can be the harm; check before you spread.*)
- A voice note that sounds exactly like your mum asks you to do something odd. What's the move? (*Verify on another channel. The voice is no longer the proof.*)

If it stalls

- **The class aces the image rounds.** Good — switch to audio: *"Your eyes are trained. Your ears aren't."* Then play the reserve pair.
- **A link is dead or the screen fails.** Run Round 0 and go straight to the debrief. The zero-technology round carries the teaching point on its own.
- **A student goes nihilist — "so nothing's real, why care?"** Don't argue — hand it to the room: *"Is that right? Is nothing real?"* Someone will push back; back them up. Plenty is real, and checking sources is how you find it — journalists, courts and scientists earn trust that way for a living. The skill is trust you can check, not doubting everything.
- **Someone asks "can we make one?"** The answer is no, and this is where the SHOW-never-make rule earns its keep. Making a fake of a real person is a consent violation, the law around it is moving fast, and the skill that will actually serve them is spotting and checking, not producing. (The full session covers what the law is starting to say, at an age-appropriate level, kept current.)

What the full session adds (honestly)

This challenge is real value, and you can run it forever without us. Here's what changes when we're in the room:

In this pack	In the full session
A challenge you source yourself	A pre-vetted, escalating deck we build and keep current – including the synthetic-audio rounds, which are the hardest to source well – with the rounds ordered so the confidence drop lands on schedule
Hallucination <i>described</i> , via reporting	Hallucination <i>demonstrated live</i> on the presenter's own machine – the room watches a confident wrong answer happen
The discussion of trust	The BUILD IT moment: your students write an AI assistant's guardrails on the big screen, try to break them, and watch one attempt get through. They learn "guardrails are just words" by breaking some
The integrity line, stated	The integrity conversation, had properly: why detectors don't work, where the line between help and cheating actually sits, and what using AI well looks like for <i>their</i> assessments
The law, mentioned carefully	The deepfake-and-consent law step, delivered by a UWA JD (law) student – current, credible, age-appropriate
One teacher running one activity	Two presenters – a working AI practitioner and a registered teacher – running all five steps of the session
This pack	A teacher resource pack you keep: pre-session primer, teacher-facing run sheet, framed video links, post-session activities and the curriculum mapping

Where it sits in the curriculum

An incursion **connects to and enriches** the curriculum – it doesn't deliver a descriptor to the achievement standard on its own, and we won't claim it does. For Years 7–8 the series speaks most directly to (Australian Curriculum v9, Digital Technologies):

Code	Descriptor	How this series connects
AC9TDI8K03	investigate how digital systems represent text, image and audio data using integers	Images and voices are just data – which is exactly why they can be synthesised, and exactly why "it looks real" proves nothing.
AC9TDI8P14	investigate and manage the digital footprint existing systems and student solutions collect, and assess if the data is essential to their purpose	The digital-citizenship step: what you make, share and feed into systems, and what it costs you.
AC9TDI8P10	evaluate existing and student solutions against the design criteria, user stories and possible future impact	Judging what's fair to make and share, and what it does to real people.

The heavier lifting sits in the **general capabilities** at this band (Level 5): **Digital Literacy** (finding and judging information – telling credible from unreliable sources is the deepfake skill), **Ethical Understanding**, and **Critical and Creative Thinking**. The series is also built to be consistent with the

Australian Framework for Generative AI in Schools — particularly its *Human and Social Wellbeing*, *Transparency*, and *Privacy, Security and Safety* principles.

The full descriptor-by-descriptor mapping lives in the C4Sight Curriculum Alignment Matrix, which every student-facing booking receives.

Where we stand on AI

You are about to put this in front of children, so you are entitled to know what the people who wrote it actually think.

We are not AI evangelists. Caleb uses these tools most working days and still would not call himself a fan. It is a drill: you pick it up when it suits the job and put it down when it doesn't. The moment it starts doing your thinking instead of your work, it has stopped being useful. That is the view underneath every one of these packs, and it is why the activities are built to make students sceptical rather than impressed.

Two things we think schools should keep well away from it: **creativity and judgement**. Drafting is labour and a machine can carry some of it. Deciding is different, and a student who never sits in the not-knowing does not build the muscle that hard problems need later. The struggle was the point of the lesson.

What that means for a C4Sight session, without exception:

- No student devices. Everything hands-on is unplugged, or driven by the presenter.
- No student data is entered into any system, ever.
- No student images are captured, saved or shown. The camera frames only the objects students hold up.
- A school staff member is present throughout.
- The tool runs on the presenter's own device and hotspot, never your network, and the classifier runs fully offline.

And the thing schools least enjoy hearing from us: AI detectors do not reliably work. They flag honest work as cheating and wave real AI writing straight through, and the students who cop it worst are often the ones who speak English as a second language. We will not sell you one, and we will say so in front of your staff. What does work is changing how the task is set, and that is teachable.

All of it is published, dated, and open to being checked: **c4studios.com.au/how-we-use-ai**

If your school would rather keep AI outside the fence entirely, that is a fair position and we will not argue you out of it. We would rather you knew where we stood before you booked us than after.

Bring the full series to your school

All prices exclude GST · Perth metro · two presenters unless noted.

Format	What it is	Reaches	Price
Single Year Group	2 × 60 min sessions (one presenter)	Up to 60	\$950
Half Day	3 sessions	Up to 90	\$1,750
Full Day	5 sessions + a free 45-min after-school staff briefing	Up to 150	\$2,850
Term Program	6 weekly sessions with one year group, ending in a student showcase	One year group	\$7,500

Single Year Group, Half Day and Full Day bookings include the teacher resource pack. Many schools pair a student day with **Staff PD (\$2,200, or \$3,400 extended)** — it comes out of a different budget line, and the integrity conversation works best when staff and students hear the same honest story.

Or start free

Book the free 30-minute Taster. One class or a short assembly, one presenter, no cost. The rules: one per school, ever; a decision-maker (Principal, Deputy, HOLA or Digital Technologies coordinator) in the room; and it ends with a scheduled conversation about where AI fits at your school.

→ caleb@c4studios.com.au · c4studios.com.au

Who you'd be booking. You don't have to take our word for anything. Things you can check yourself: we run a Perth studio building AI automation for commercial clients — real, current work. Our practitioner is a UWA JD (law) student, so the ethics, consent and regulation content is genuine, not borrowed. We produce the C4Sight explainer series — public work you can watch before you book. And we deliver alongside a TRBWA-registered teacher.

C4Sight is a C4 Studios venture · Caleb Scott · ABN 26 477 523 297 · Perth, WA
