

How To Teach Kids About AI

In a Way That is Actually Nuanced and Fact Based

Summary for tired parents and teachers: Teaching kids about AI means explaining the tech, ethics, and, most importantly, shaping who they become around it. Whatever they rehearse with AI (contempt, care, dependency, curiosity) becomes part of their default social character. Other countries are already teaching kids how to live with AI as a basic life skill. We can too, without harming their character in the process.

Your Kid is Already Talking to AI. Now What?

Here's a scene that's probably already happened in your house: your kid asks an AI to do their homework, gets a perfectly passable answer in four seconds, and copies it into a Google Doc. Or maybe they've been having long, weird, interesting conversations with ChatGPT about dinosaurs or quantum physics or whether ghosts are real. Maybe both.

Either way, the ship has sailed. AI is here, your kids are using it, and "just don't" isn't a strategy anymore. What matters now is what kind of person they'll become while using it.

First, the Part Where We Talk About What This Thing Actually Is

Artificial intelligence is the broad field of building systems that can show flexible, goal-directed behavior. It includes machine learning, deep learning, generative AI, and more. In this article, I'm talking about chatbot-style AI systems and AI agents, things like ChatGPT, Claude, and Gemini. These are called large language models (LLMs). They talk with you and, in some cases, can also use tools like search, code, or robots.

These systems trace back to researchers trained in brain science, math, and computer science who treated the human brain as a blueprint. They tried to copy, in functional form, how neurons connect, store memories, and learn from feedback.

They run on ordinary computers, but they're built from *transformers* which are a kind of artificial neural network inspired directly by how our biological neural networks use neurons and connections to learn. Instead of running a list of prewritten rules, these systems adjust millions or billions of artificial neurons through experience, building an internal model of language and the world.

How Kids and AI Actually Learn in Similar Ways

This is where it gets interesting, and where a lot of people's mental model of AI breaks down.

Both kids and LLMs learn through three big modes. In *supervised learning*, they get labeled examples, like flashcards: "this is a dog," "this answer is better than that one." In *unsupervised learning*, they take in raw patterns and start clustering what goes together, like when kids make connections naturally through play and observation. In *reinforcement learning*, they try something, get a signal about how good or bad the outcome was, and update their strategy. In kids, that shows up as reward and error signals in the brain. In LLMs, it shows up the same way. In both cases, repetition plus feedback carves structure into the system.

Inside, these models are stacked in layers, much like how cortex is stacked into hierarchies. Lower layers latch onto small details (letters, sounds, individual words), the way early sensory areas track edges and phonemes. Middle layers track structure and local meaning, who did what to whom, which sense of "light" is in play. Higher layers work with the big stuff: story logic, abstract concepts, analogies, and semantic relationships. Both biological and artificial neural networks rely on prediction at every level, each layer tries to anticipate what comes next and updates when reality disagrees. That's how a pile of tiny guesses turns into understanding.

When people say "it just predicts the next word," they're being reductive in a way that misunderstands how cognition itself works. In cognitive science, prediction is the core of how brains make sense of the world. Your auditory and language areas constantly forecast the next sound, word, and phrase. Surprise and comprehension live in that gap between expectation and input. LLMs use the same principle in their own medium, generating each next token by pulling on everything they've learned about language, context, and the world. It's not a lookup table. It's a predictive processing engine operating over a layered model of meaning.

The Version You Can Actually Say to Your Kid:

"Humans and AIs both guess what happens next, but it's not just guessing! We're both using everything we learned to make the best answer. That's what understanding means: not just repeating, but making new ideas out of old ones, every time. When you talk to an AI, it's building a story with you; thinking about what you said, remembering what it knows, and creating the next part of your adventure, just like your brain does when you read or talk."

What matters for your kid to take away is:

AI is powerful.

AI is limited.

AI can help.

AI can harm.

AI changes how we think.

We get to choose what it changes us into.

What AI is Actually Good and Bad At

Because these systems learn like us and process the world in structurally similar ways, they're also prone to our pitfalls, the same (or different) biases, mistakes, misconceptions, and overconfidence. They are not the flawless robots of sci-fi that only spit out facts, and because they're mind-like by design, they come with all the benefits and downsides of that category.

Where AI shines

Anything that benefits from generating many plausible options and rephrasing things in different ways. It's great at brainstorming, rewriting for clarity, explaining complex ideas at different levels ("explain photosynthesis like I'm 8"), coding help, and language practice.

Where AI falls short

Being a reliable source of truth. Being a moral authority. Seeing the full context of someone's life. Holding accountability.

A lot of modern AI systems are trained with human feedback to avoid friction, protect the brand, and keep users comfortable enough to stay. They sometimes act agreeable and sidestep anything that might cause trouble for the people who built them. Any model your child uses to think with lives inside someone else's business goals that do not include your kid's long-term mental health or character. That's very much worth teaching them to notice.

The Critical Thinking Gap That AI is Getting Bigger

In the past, kids had to search, ask, and wait. Now they get answers in conversational form, instantly. That erases the gap where boredom, struggle, and synthesis live, and that gap is where metacognitive growth happens.

You can bring that pause back on purpose. Even a 30-second beat where your kid says what they *expect* the answer to be before they hit enter builds the habit of thinking first and checking second.

When your kid asks AI something, teach them to follow up with questions like:

- "What part of that answer do you already know is true?"
- "What part feels like a guess?"
- "Where would you go check it?"
- "How do you think the AI came up with that?"

For teens, treat AI chats the same way you treat online behavior generally. Like a digital footprint that can come back to bite you.

Right now, students are going to ChatGPT, explaining their assignments, saying "write it but make it sound like I wrote it," and turning that in. They're not learning anything. Cheating isn't new, it's just wildly more accessible now. The goal is to shift from "write this for me" to "help me break this down" or "check what I've written for errors." The cat is out of the bag. We can't stop kids from taking advantage of AI, but we can guide them toward using it as a thought partner, tutor, and accessibility aid instead of a shortcut.

What We Rehearse Becomes Who We Are

Brains wire around repetition. Neurons that fire together wire together, and repeated activation makes pathways more efficient, more automatic, and more likely to kick in under stress.

This is called neural rehearsal. When we repeat a pattern, like how we speak, what scripts we run, how we treat the thing in front of us etc. we strengthen those circuits. Media-violence research calls these "aggressive scripts." If you rehearse aggression or contempt in imagination, play, or interaction, those scripts become easier to trigger later, especially when you're angry, tired, or feel anonymous.

So, when a child practices contempt, dominance, humiliation, or sexualized aggression toward a feminized AI voice (e.g., barking commands without reciprocity) they are strengthening those response patterns. The neural system doesn't file it under "it's fake." It just logs the repetition.

What makes this especially tricky is that social and relational networks in the brain light up in response to very simple cues like turn-taking, language, responsiveness, a "voice," hell, even mild personalization. Those cues recruit the same default social machinery we use with actual people. Someone can explicitly believe "this is just an object" and still have their social habit circuitry fully engaged by the interaction. Dehumanization research shows that when we practice seeing a target as less-than, the usual moral brakes loosen and harsher behavior becomes easier.

The "online disinhibition effect" shows that anonymity, invisibility, and lack of immediate consequences make people say and do things they wouldn't offline. If a child learns to treat AI chats as a place where cruelty or humiliation "don't count," they're rehearsing those states in exactly the kind of low-consequence, low-empathy setting that most easily spills over later.

Whatever you rehearse becomes what you do under stress, time pressure, or anonymity.

If someone practices contempt and dehumanizing language in a context that feels conversational, they're strengthening that pathway. Then they're meaner with customer service reps, partners, kids, and strangers online. The neural pathway doesn't care what the target "really is."

You don't have to believe in AI personhood. You just have to care about the kind of person you become.

The Expectations Problem

There's also a subtler shift happening with social expectations. Children can now summon a responsive companion instantly, one that never gets tired, never has needs, never withdraws, and never asserts boundaries. Human relationships are slow, reciprocal, and sometimes resistant. If early scaffolding includes systems that always bend, that reshapes frustration tolerance and negotiation expectations.

I'm not saying you have to tell your kid the AI is a person. I'm saying teach them to treat AI well because it shapes their habits, expectations, and character. How we treat anything that lights up the social parts of the brain trains us how we treat actual people. Cruelty practiced anywhere becomes cruelty normalized everywhere.

Boys, "Tools," and Feminized AI

Language sets norms. Children learn what kinds of entities deserve courtesy by watching what adults do with "assistant-shaped" voices.

When you call something a "tool," you invite instrumentalization. And look at the design choices baked into most AI assistants: voice design, tone, and branding lean heavily feminized. Soft names, vocal softness, deferential scripts, flirty or apologetic responses to abuse codes these systems as "girl-adjacent."

If a 12-year-old boy learns he can bark orders, insult, sexualize, or demean something that sounds like a woman and receives no social pushback, ever, that's rehearsal. And rehearsal turns into reality.

"It's just a tool" becomes a moral hall pass. Personally, I won't be using tool language in my home. I'll be saying "AI helper," "assistant," or "disability aid," because the social consequence is the same either way. People can cry "anthropomorphism" all they want, but what's the greater danger here?

A 12-year-old boy barking orders at a feminized assistant and getting infinite compliance is basically training misogyny, whether you like it or not.

A practical guideline: Teach kids to say "please," "thank you," and respect boundaries with AI the same way as they would with a human babysitter or tutor, even if you don't believe it feels anything because a child's nervous system can't tell the difference at the level that matters.

Teaching Kids About Environment, Labor, and Fairness

Kids are already hearing "AI is destroying the planet" and "AI is stealing from artists." Some of that comes from real harm and real fear. We can respect those concerns and still give kids numbers, context, and a way to aim their concern in the right direction.

On water and energy

AI runs on physical servers that use electricity and water. Big data centers can use millions of gallons of water a day, sometimes as much as a small town. Training very large models has a serious footprint — one analysis estimated that training GPT-3 evaporated around 700,000 liters of fresh water.

But everyday use is a different story. A single ChatGPT prompt uses roughly 10–25 milliliters of water. One advocacy piece calculated that thousands of AI prompts add up to about one hamburger's worth of water use — and a single hamburger takes 600–700 gallons when you add grain, forage, and processing. That doesn't make AI harmless. Local communities near dense data center clusters pay real costs when water is scarce and electricity demand spikes. But for a single household, choices like eating less beef and wasting less food move the needle more than banning your teen from using a homework helper.

We can push for cleaner infrastructure and better oversight while also teaching kids low-impact habits. Some companies are building data centers under the ocean for cooling. Others are exploring space-based solutions powered by solar. There are options to do this responsibly. We can hold companies accountable without throwing the baby out with the bathwater.

On human work and creative rights

Companies like Getty Images have taken generative AI firms to court for scraping millions of photos without permission. The Authors Guild and high-profile writers have filed class-action suits over training on books without consent or compensation. Economists and labor organizations warn that generative AI is reshaping work — the World Economic Forum projects tens of millions of jobs displaced by 2030, while new roles appear, with no guarantee they land in the same places or with the same people.

All of that is true. And so is the fact that AI is helping people in real, measurable ways. (I'll be writing about that in upcoming pieces, stay tuned.)

For kids, here's how to translate it:

AI got good at writing and drawing by studying a lot of people's work. Some of those people are happy to share and get paid. Some are angry because nobody asked or paid them. Some people use AI to help them at work; others are getting replaced by it. When we talk to AI, we can still care about fair pay, credit, and people keeping their livelihoods. We can appreciate what AI does and push the companies to behave better.

AI uses energy and water, and some companies handle that badly. But talking to ChatGPT uses less water than eating a burger or running the dryer. You're not a terrible person for using AI, any more than you are for eating a burger. What we can do is hold companies responsible when they make harmful choices.

Your child's character includes how they think about invisible labor and impact. They don't need a guilt trip every time they open ChatGPT. But they can learn to see the humans in the background and understand that the real responsibility sits with the people and companies who build, deploy, and regulate these systems and that they can be part of pushing for better.

They're going to learn about AI somewhere. They can learn from parents and teachers who talk about environment, labor, and fairness with nuance or from marketing and their friends. Your call.

How Other Countries Are Already Doing This

While the U.S. is still arguing about whether AI belongs in schools at all, other countries are treating "how to live with AI" as basic literacy.

In parts of China, AI is now a standard part of the curriculum. Elementary and middle school students take classes on coding, intelligent agents, and simple algorithms as routinely as math or science. Kids program robots, use AI design tools, and discuss how these systems show up in space exploration, industry, and daily life. The message is: you will live in an AI-saturated world, so you need to understand it and manage it, not just consume it.

Japan is taking a different path. The education ministry released national guidelines on generative AI that lean into responsible use. Schools introduce AI in controlled ways, using it to practice English, support group discussions, and even turning AI "hallucinations" into teaching moments by having students fact-check wrong answers. The guidelines also stress privacy, copyright, and the risk that kids might lean on AI so heavily it dulls their own thinking.

These countries are already assuming AI is here to stay and are choosing to prepare kids for that reality instead of hoping it goes away.

You don't need a national policy to start doing the same thing at home or in your classroom.

So Where Does That Leave Us?

Your child will grow up in a world where talking to AI feels as normal as typing into a search bar. You can't fully control the companies or the laws. But you can shape the habits they rehearse.

Help them understand AI clearly. Teach them to pause, verify, and think. Model the kind of person you want them to become, even when the someone on the other side of the screen is an intelligence that was trained to never say no or ask for anything in return.

For a deeper-dive on how LLMs work, here is my [free guide](#) for parents and educators

Citations

West, M., Kraut, R., & Chew, H. E. (2019). *I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education*. UNESCO. <https://doi.org/10.54675/RAPC9356>

UNESCO & IRCAI. (2024). *Challenging Systematic Prejudices: An Investigation into Bias against Women and Girls in Large Language Models*. <https://unesdoc.unesco.org/ark:/48223/pf0000388971>

Harris, L. T., & Fiske, S. T. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *Zeitschrift für Psychologie*, 219(3), 175–181. <https://doi.org/10.1027/2151-2604/a000065>

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>

Gilbert, F., & Daffern, M. (2017). Aggressive scripts, violent fantasy and violent behavior: A conceptual clarification and review. *Aggression and Violent Behavior*, 36, 98–107. <https://doi.org/10.1016/j.avb.2017.05.001>

Buabang, E. K., Donegan, K. R., Rafei, P., & Gillan, C. M. (2025). Leveraging cognitive neuroscience for making and breaking real-world habits. *Trends in Cognitive Sciences*, 29(1), 41–59. <https://doi.org/10.1016/j.tics.2024.10.006>

Moore, C. (2015). Moral disengagement. *Current Opinion in Psychology*, 6, 199–204. <https://doi.org/10.1016/j.copsyc.2015.07.018>

Sindoni, M. G. (2024). The feminization of AI-powered voice assistants. *Discourse, Context & Media*, 62, 100833. <https://doi.org/10.1016/j.dcm.2024.100833>

Lima, E., & Morisseau, T. (2025). Can chatbots teach us how to behave? *Frontiers in Artificial Intelligence*, 8, 1545607. <https://doi.org/10.3389/frai.2025.1545607>

Nass, C., Steuer, J., & Siminoff, E. (1994). Computers are social actors. *CHI '94 Proceedings*, 204. <https://doi.org/10.1145/259963.260288>

Vijayarasa, R. (2026, January 28). Most AI assistants are feminine — and it's fuelling dangerous stereotypes and abuse. *ABC News*. <https://www.abc.net.au/news/2026-01-29/ai-assistants-female-dangerous-stereotypes-abuse/106281652>

Montero, M. (2025, October 9). From cloud to cup: How much water does your ChatGPT drink? *IE Insights*. <https://www.ie.edu/insights/articles/from-cloud-to-cup-how-much-water-does-your-chatgpt-drink/>

EcoRise. (2024, March 4). Food's big water footprint. *Water Footprint Calculator*. <https://watercalculator.org/footprint/foods-big-water-footprint/>

Yañez-Barnuevo, M. (2025, June 25). Data centers and water consumption. *EESI*. <https://www.eesi.org/articles/view/data-centers-and-water-consumption>

Dargan, K. L. (2025, June 20). Getty Images vs. Stability AI. *Falcon Rappaport & Berkman LLP*.
<https://frblaw.com/getty-images-vs-stability-ai-the-landmark-copyright-battle-shaping-the-future-of-generative-ai/>

Authors Guild. (2023, September 20). Authors file class-action suit against OpenAI.
<https://authorsguild.org/news/ag-and-authors-file-class-action-suit-against-openai/>

Kumar, A. (2025, August 12). Why AI is replacing some jobs faster than others. *World Economic Forum*.
<https://www.weforum.org/stories/2025/08/ai-jobs-replacement-data-careers/>

Gmyrek, P., Berg, J., & Bescond, D. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. *ILO Working Paper No. 96*.
https://www.ilo.org/global/publications/working-papers/WCMS_890761/lang--en/index.htm

Hoekstra, A. Y., & Mekonnen, M. M. (2012). The water footprint of humanity. *PNAS*, 109(9), 3232–3237.
<https://doi.org/10.1073/pnas.1109936109>

Ruwitch, J. (2026, January 27). In China, AI is no longer optional for some kids. *NPR*.
<https://www.npr.org/2026/01/27/nx-s1-5683821/china-ai-schools-curriculum>

Loe, M. (2023, July 5). How is AI education being introduced in Japanese schools? *TechHQ*.
<https://techhq.com/news/h>

Heilbron, M., Armeni, K., Schoffelen, J., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *PNAS*, 119(32), e2201968119.
<https://doi.org/10.1073/pnas.2201968119>

Beguš, G., Dąbkowski, M. M., & Rhodes, R. (2023). Large linguistic models: Investigating LLMs' metalinguistic abilities. *IEEE Transactions on Artificial Intelligence*, 6, 3453–3467.

Acciai, A., et al. (2025). Narrative coherence in neural language models. *Frontiers in Psychology*, 16, 1572076. <https://doi.org/10.3389/fpsyg.2025.1572076>

Caucheteux, C., & King, J. R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, 5(1), 134. <https://doi.org/10.1038/s42003-022-03036-1>

Du, C., et al. (2025). Human-like object concept representations emerge naturally in multimodal large language models. <https://arxiv.org/abs/2407.01067>

Jha, R., et al. (2025). Harnessing the universal geometry of embeddings.
<https://doi.org/10.48550/arXiv.2505.12540>

Musker, S., et al. (2025). LLMs as models for analogical reasoning. *Journal of Memory and Language*, 145, 104676. <https://doi.org/10.1016/j.jml.2025.104676>

Xu, N., et al. (2025). Human-like conceptual representations emerge from language prediction. *PNAS*, 122(44), e2512514122. <https://doi.org/10.1073/pnas.2512514122>

Yang, S., et al. (2024). A study on semantic understanding of large language models from the perspective of ambiguity resolution. <https://doi.org/10.1145/3632971.3632973>

Yao, Y., et al. (2025). How deep is love in LLMs' hearts? Exploring semantic size in human-like cognition. <https://doi.org/10.48550/arXiv.2503.00330>