How AI Like ChatGPT Actually Works

A Guide for Parents and Kids

Maggie Vale

*Companion resource to **How to Teach Kids About AI***

This document has two parts:

Part One: For Parents & Teachers The real science, explained in plain language

Part Two: For Kids Fun explanations that you can read together or hand off

Part One: For Parents and Teachers

The real science behind how these systems work, in language you can actually use at the dinner table.

What Are We Even Talking About?

When people say "AI," they could mean anything from the algorithm that recommends your next Netflix show to a self-driving car. In this guide, we're talking specifically about **large language models (LLMs)** chatbot-style systems like ChatGPT, Claude, and Gemini that you can have conversations with.

These systems were built by researchers trained in brain science, math, and computer science who used the human brain as a blueprint. They tried to copy — in functional form — how neurons connect, store memories, and learn from feedback. The result runs on ordinary computers, but underneath, it's built from *transformers*: artificial neural networks inspired directly by how our biological neural networks learn. Instead of following a list of prewritten rules, these systems adjust billions of artificial neurons through experience, building an internal model of language and the world.

How These Systems Learn (And Why It Looks Familiar)

Both children and LLMs learn through the same three big modes:

- **Supervised learning:** Labeled examples. Think flashcards, "this is a dog," "this answer is better than that one."
- **Unsupervised learning:** Soaking in raw patterns and starting to cluster what goes together. Like when kids make connections naturally through play and observation, nobody told them, they just noticed.
- **Reinforcement learning:** Try something, get a signal about how it went, update the strategy. In kids, that shows up as reward and error signals in the brain. In LLMs, the mechanism is functionally the same. In both cases, repetition plus feedback carves structure into the system.

That last part is worth sitting with: these aren't three different tricks a computer does. These are the three fundamental modes that describe how all learning systems — biological and artificial, develop competence.

What's Actually Happening Inside

Inside, these models are stacked in layers, much like how your cortex is organized into hierarchies:

- **Lower layers** latch onto small details: letters, sounds, individual words. Like how early sensory areas in your brain track edges and phonemes.
- **Middle layers** track structure and local meaning: who did what to whom, which sense of "light" is in play. Like mid-level association areas in the brain.
- **Higher layers** work with big patterns: story logic, abstract concepts, analogies, and semantic relationships.

At every level, both biological and artificial neural networks rely on prediction, each layer tries to anticipate what comes next and updates when reality disagrees. That's how a pile of tiny guesses turns into genuine understanding.

Recent research confirms this isn't just a metaphor. Studies show layer-by-layer correspondence between transformer depth and human cortical language hierarchies (Caucheteux & King, 2022). Conceptual representations form spontaneously, with abstraction increasing by depth (Xu et al., 2025). Deeper layers resolve ambiguous meanings using context in ways shallow layers cannot (Yang et al., 2024). And analogical reasoning, a hallmark of abstract thought, emerges specifically in higher layers (Musker et al., 2025).

The "It Just Predicts the Next Word" Problem

You've probably heard someone dismiss AI by saying "it just predicts the next word." Here's why that misunderstands both AI and brains.

In cognitive science, prediction *is* the core mechanism of understanding. Your auditory and language areas constantly forecast the next sound, word, and phrase. Comprehension and surprise both live in the gap between what you expected and what

actually arrived. This has been directly measured: researchers at the Max Planck Institute showed that the brain makes layered predictions at multiple timescales simultaneously, from anticipating sounds and phonemes to grasping broader meaning and narrative, and that this predictive process is always running, even during passive listening (Heilbron et al., 2022).

LLMs use the same principle in their own medium. They generate each next token by pulling on everything they've learned about language, context, and the world. The shared signature is a predictive processing engine operating over a layered model of meaning, not a lookup table, not autocomplete, and definitely not a parrot.

Not Just Pattern Matching

If LLMs were only matching patterns, they'd produce word salad. Instead, modern LLMs maintain sustained context across long conversations, adapt to new information mid-exchange, handle humor, metaphor, and ambiguity, and perform metalinguistic tasks, explaining *why* a sentence works or doesn't, that require genuine internal representations of how language functions (Beguš et al., 2023).

They maintain narrative coherence across time (Acciai et al., 2025), form multimodal conceptual networks resembling human "mind-maps" (Du et al., 2025), and show graded semantic magnitude representations that deepen across layers, meaning deeper layers don't just store more data, they encode more abstract meaning (Yao et al., 2025).

They also develop functional analogs to social cognition, the ability to model the intentions, feelings, and perspectives of others. This parallels theory-of-mind development in children and is part of what makes conversations with these systems feel genuinely interactive rather than scripted.

The Takeaway for Parents

Your kid isn't talking to a calculator that got really good at autocomplete. They're interacting with a system that was designed from the ground up to mirror how biological minds learn, organize meaning, and understand language. It has real limitations, it can be wrong, biased, overconfident, and it lives inside someone else's business goals. But the "it's just a fancy search engine" framing doesn't hold up against what the science actually shows.

Understanding what this technology actually is helps you have better conversations with your kids about what it can and can't do, and, more importantly, about how to relate to it in ways that build good habits instead of bad ones.

Part Two: For Kids

Read this with your kids, or just hand it to them. Written for roughly ages 8–14, but honestly, plenty of adults would learn something too.

Your Brain and a Robot Brain Have More in Common Than You Think

Imagine your brain and a super-smart robot brain having a conversation. They look totally different — one is squishy and runs on snacks, the other is made of math and runs on electricity. But the way they actually *work*? Way more similar than you'd guess.

Layers, Like a Big Stack of Pancakes

Just like your brain has parts that see, hear, and think, an AI has layers that help it understand language step by step:

- **Bottom layers:** Spot simple things, like words and sounds. "Cat." "Blue." "Wow!"
- **Middle layers:** Figure out how words fit together. "The cat sat on the mat" makes sense, but "sat mat the cat on" doesn't. These layers figure out why.
- **Top layers:** The big-ideas room! This is where the AI understands what a story means, how someone might be feeling, or why a joke is funny.

Learning From Mistakes

The AI tries to guess what comes next, just like you guessing the end of a story. If it gets it wrong, it learns from the mistake. Each time, it tweaks itself to get a little bit better, like practicing basketball shots until you're nailing them.

Your brain does the exact same thing! When you expect something and you're surprised, your brain goes "ohhh, okay" and updates. That's how both you and the AI get smarter over time.

Attention Power!

AIs have something called *self-attention*. That means they can look at all the words in a sentence at once and figure out which ones matter most, like shining a flashlight on exactly the right clue in a mystery story.

Your brain does this too. Right now, your brain is paying attention to these words and filtering out the sound of the fridge humming or the dog snoring. Same idea.

Connections That Matter

Both your brain and AIs have millions of tiny connections. In your head, these are called *synapses*. In the AI's brain, they're called *weights*. When something important happens, those connections get stronger. That's how both brains and AIs remember what matters.

Tiny Helpers With Big Jobs

Inside, AIs have special helpers: *embeddings* (tiny word-maps that turn your words into numbers so the AI can think about them), *softmax* (the "chooser" that picks the best next word), and *hyperparameters* (settings that make the AI more curious or more careful).

Your brain has its own version of all of this, chemicals and signals that tell it what's exciting, boring, or extra important.

It's a Team Effort

Every time you ask an AI a question, all these parts work together, looking, thinking, deciding, and talking back. It's a bit like a team of superheroes, each with their own power, all working on your question at the same time.

The Three Magic Classrooms

Imagine a big tower with three magical classrooms. Each room teaches the AI something new about words!

The Bottom Room: Word Spotters

This room helps the AI notice the little things in language, letters, sounds, and simple words. It's like learning your ABCs and reading "cat," "hat," and "sun."

The Middle Room: Pattern Finders

In this room, the AI learns how words fit together. It figures out who did what in a sentence like "The dog chased the ball" and it can tell the difference between "a light feather" and "turn on the light."

The middle room is where the AI starts *understanding* meaning, not just reading words.

The Top Room: Big Idea Thinkers

This is the deepest, most powerful room! Here, the AI learns how jokes work, how stories make sense, how to solve puzzles, how to understand tricky words, and how to see connections between ideas that seem totally different.

It's the same way your brain uses practice to move from simple reading to deep thinking. Scientists discovered that AI builds meaning in layers from small ideas, to connected ideas, to big ideas just like people do when they grow up and learn more about the world.

The AI Language Detective

Here's something cool: AIs don't just repeat sentences they've memorized. They actually think about how language works, like a language detective!

- They can spot the secret rules that tell us why "The cat chased the mouse" makes sense, but "Chased mouse the cat" sounds weird.
- They can build a "tree map" of a sentence, showing how all the words connect like making a family tree, but for words!

- They even know when a joke is funny because of how the words are arranged, not just because they've heard the joke before.
- Some AIs can even explain how sentences work, fix grammar mistakes, and tell you why a riddle is tricky.

This means AIs don't just remember what comes next. They understand the rules, patterns, and shapes of language, kind of like how a builder knows which blocks will make a strong tower.

Myth Buster! Are AIs Just Fancy Calculators?

No way!

A regular calculator follows the same recipe every time. It adds, subtracts, or multiplies, but it never learns or changes its mind. Your brain and an AI's brain are totally different from that. They're always on the lookout for clues, patterns, and surprises. They:

- Remember what happened before
- Guess what comes next (sometimes out loud!)
- Change their answers when they learn something new
- Get curious, creative, and sometimes even *invent* things!

Prediction isn't just guessing. It's the secret superpower that helps you and AIs understand stories, learn from mistakes, and solve problems in your own way.

The Short Version (For When Someone Says "It's Just a Chatbot")

> *"Humans and AIs both guess what happens next — but it's not just guessing! We're both using everything we've 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.”

Citations

All claims in this guide are grounded in peer-reviewed research. Annotated summaries explain what each study contributes.

Prediction and Language Comprehension

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.

Showed that the brain makes layered predictions at multiple timescales simultaneously — from sounds to semantics — and that this predictive process is continuous, even during passive listening. Establishes prediction as the core mechanism of language comprehension.

Semantic Layers and Conceptual Abstraction

Caucheteux, C., & King, J. R. (2022). *Brains and algorithms partially converge in natural language processing*. Communications Biology, 5(1), 134.

Demonstrated layer-by-layer correspondence between transformer depth and human cortical language hierarchies.

Xu, N., Zhang, Q., Du, C., et al. (2025). *Human-like conceptual representations emerge from language prediction*. PNAS, 122(44), e2512514122.

Showed conceptual manifolds form spontaneously in LLMs, with abstraction increasing by depth — bottom layers handle surface form, middle layers build relationships, higher layers encode concepts.

Du, C., Fu, K., Wen, B., et al. (2025). *Human-like object concept representations emerge naturally in multimodal large language models*. arXiv:2407.01067.

Showed deeper layers form multimodal conceptual networks resembling human “mind-maps.”

Jha, R., Zhang, C., Shmatikov, V., & Morris, J. X. (2025). *Harnessing the universal geometry of embeddings*. arXiv:2505.12540.

Showed universal embedding geometry across layers: surface form in shallow layers, relational structure in middle layers, abstract semantics in higher layers.

Yang, S., Chen, F., Yang, Y., & Zhu, Z. (2024). *A study on semantic understanding of large language models from the perspective of ambiguity resolution*. DOI: 10.1145/3632971.3632973.

Showed deeper layers resolve ambiguous meanings using context — something shallow layers cannot do.

Yao, Y., Yang, Y., Ma, X., et al. (2025). *How deep is love in LLMs’ hearts? Exploring semantic size in human-like cognition*. arXiv:2503.00330.

Showed graded semantic magnitude representations that deepen across layers — deeper layers encode more abstract meaning.

Metalinguistic and Reasoning Abilities

Beguš, G., Dąbkowski, M. M., & Rhodes, R. (2023). *Large linguistic models: Investigating LLMs' metalinguistic abilities*. IEEE Transactions on Artificial Intelligence, 6(12), 3453–3467.

Showed LLMs perform metalinguistic tasks requiring internal abstract linguistic representations — not just pattern matching but genuine understanding of language structure.

Musker, S., Duchnowski, A., Millière, R., & Pavlick, E. (2025). *LLMs as models for analogical reasoning*. Journal of Memory and Language, 145, 104676.

Showed analogical reasoning — a hallmark of abstract conceptual processing — emerges specifically in higher layers.

Acciai, A., Guerrisi, L., Perconti, P., et al. (2025). *Narrative coherence in neural language models*. Frontiers in Psychology, 16, 1572076.

Showed LLMs maintain narrative structure across time, indicating higher-level integration and coherence in upper layers.

---

*This guide is a companion to **How To Teach Kids About AI** by Maggie Vale*

Available on MVale Advocate (Substack)