AI Explained

Knowledge, Search and Tools (How AI Knows Things)

By Dr Peter Fields

THE 2-MINUTE SUMMARY

An AI answer can draw on patterns learnt during training, information you supply, or data returned by a tool. A model's published knowledge cutoff is a rough boundary for its built-in knowledge. It tells you little about whether a particular book, rule or event is adequately represented.[1,2,3]

Search gives an assistant access to fresh evidence. The answer still depends on which sources it finds, what material it obtains and whether that material supports its claims. A citation is a place to check.[3,4,5]

Models can also work with images and audio. Some use image patches or audio representations. The details vary by system, and a model that describes the overall scene accurately can still misread a small detail.[6,8,9,10]

The contradiction on the screen

Ask who won last night's Premier League match and the assistant may give you the exact score. Ask about a novel from eight months ago and it may invent a plot. Then upload a street photograph. The assistant might describe the light and architecture convincingly, yet misread the church clock. The score, the plot and the clock each ask the system to obtain and use a different kind of evidence.

Where an answer gets its information

A language model's weights are numerical parameters adjusted during training. They encode patterns that can support factual answers, among many other capabilities. The weights are a poor substitute for an indexed catalogue of what the model has seen. A fact can pre-date the published cutoff and still be absent from an answer or recalled incorrectly. Different model versions can have different cutoffs.[1,2]

An assistant may also be given the recent score in your message, a contract in an upload, or a private document through a connected search tool. The surrounding application can decide which tools to make available and whether to invoke them directly or allow the model to request them. Retrieval can target a collection you supply as well as the public web. In the opening example, a search or live feed could supply the score. If the assistant tries to recall the plot from training alone, it may have too little reliable information about the book to answer.[1,2,3,4]

Geek Box · What happens when a tool is called?

A tool call can be a structured request. For a weather query, the model could supply the location and date in separate fields. The application receives the request, runs the tool if appropriate and makes the returned data available for the next step. A calculator can return a number and a search tool can return links or extracted passages. The exact route varies. Toolformer demonstrated one training method for using such tools. WebGPT investigated one approach to browsing and citing sources.[3,4]

Why a live answer can still be wrong

Search can retrieve an obsolete rule, a report about a different company with the same name, or a headline without the qualification in the article. Using a good source also requires reading it correctly. In 2023, Liu and colleagues audited Bing Chat, NeevaAI, Perplexity.ai and YouChat. On average, citations fully supported 51.5% of the sentences the engines generated. Of the citations provided, 74.5% supported the sentence they accompanied. The two figures answer different questions about the systems tested at the time.[5]

Suppose an answer says a venue opens on Sunday mornings and links to its website. The link may lead to ordinary opening hours, while the relevant page carries a bank-holiday exception. Find the passage that answers the question for the actual date. The same habit applies to a changing regulation or the latest version of a company policy.

How AI works with images and audio

Imagine cutting a picture into square tiles and recording where each tile sat. The Vision Transformer does something similar, turning the tiles into a numbered sequence while keeping their positions. A text token is a small unit of text, such as a word, part of a word or a punctuation mark. In this vision model, each unit represents an image tile. CLIP shows another way of working with images and language. It learnt to match images with their text descriptions.[6,7]

Audio can also be processed in several ways. For speech recognition, Whisper makes a kind of picture of the sound, showing how its frequencies change over time. A product can pass a transcript to a language model. Other systems can process audio as part of an end-to-end model. A useful distinction is whether the assistant received words from a transcript, audio from which it might use additional cues, or both.[8]

> **Geek Box · Why the street scene is easier than the clock**
>
> A street description can draw on many visible cues. The time on a small clock may depend on a few pixels. Image resizing or an indistinct source can make those details harder to read. Controlled studies have also found failures on simple visual tasks with clear images, so there is more to the problem than resolution alone. In a 2025 study of seven models reading analogue clocks, the best scored about 23% on exact matches, and six scored below 10%. Check the dial itself if the exact time matters.[9,10,11]

How to use it better

- **Trace a fresh claim to its source.** For last night's score, look for a search indicator or citation and check the match date against the results page. Open the cited page, read the relevant passage and check the date, edition or jurisdiction.
- **Give the assistant the source you want used.** For private or recent material, supply the relevant extract and ask it to identify the passage supporting each conclusion. Check the extract and any qualifications.
- **Provide an overview and a close-up if the detail matters.** The overview preserves context. The close-up can help with small text or a dial. Check precise readings and counts against the original.

Read more

Hallucinations, Part 4, **It Has Sources. Can You Trust It Now?**, looks more closely at what a citation can establish.

Sources and notes

1. Shanahan (2024), Talking about Large Language Models, Communications of the ACM 67(2), pp 68–79. Background on pretrained models and systems built around them. Source
2. OpenAI, Models, API documentation. Model pages provide examples of separately documented knowledge cutoffs. Source
3. Schick et al. (2023), Toolformer: Language Models Can Teach Themselves to Use Tools, NeurIPS, pp 68539–68551; Google, Function Calling with the Gemini API. A research example and a documented tool-calling workflow. Source 1 · Source 2
4. Nakano et al. (2021), WebGPT: Browser-Assisted Question-Answering with Human Feedback, arXiv preprint; Lewis et al. (2020), Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, NeurIPS. Source 1 · Source 2
5. Liu, Zhang and Liang (2023), Evaluating Verifiability in Generative Search Engines, Findings of EMNLP, pp 7001–7025. Historical audit of citation recall and precision. Source
6. Dosovitskiy et al. (2021), An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale, ICLR. See §3.1 on patches and position embeddings. Source
7. Radford et al. (2021), Learning Transferable Visual Models from Natural Language Supervision, ICML, PMLR 139, pp 8748–8763. The CLIP paper. Source
8. Radford et al. (2023), Robust Speech Recognition via Large-Scale Weak Supervision, ICML, PMLR 202, pp 28492–28518; OpenAI (2024), GPT-4o System Card. Source 1 · Source 2
9. Tong et al. (2024), Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs, CVPR, pp 9568–9578; Rahmanzadehgervi et al. (2024), Vision Language Models Are Blind, ACCV. Source 1 · Source 2
10. OpenAI, Images and Vision, API documentation. Examples of documented limitations on small text, spatial localisation and counting. Source
11. Saxena, Gema and Minervini (2025), Lost in Time: Clock and Calendar Understanding Challenges in Multimodal LLMs, ICLR Workshop on Reasoning and Planning for LLMs, Table 1. Source