"It Does Not Look Things Up": The One Idea That Demystifies AI

Hafiz Hanif

Faculty of Human Development, Sultan Idris Education University · ORCID 0000-0001-9943-6557

CITE AS Hanif, H. (2026, August 19). "It Does Not Look Things Up": The One Idea That Demystifies AI. Hack(edu).tech Chronicle. https://hackedu.tech/articles/it-does-not-look-things-up-the-one-idea-that-demystifies-ai

PUBLISHED 19 August 2026

DOI 10.5281/zenodo.22005873

LICENCE CC-BY-4.0

AVAILABLE AT https://hackedu.tech/articles/it-does-not-look-things-up-the-one-idea-that-demystifies-ai

ABSTRACT

Public understanding of AI oscillates between awe and dismissal. This article proposes a single demystifying principle: the AI model predicts plausible continuations; it does not look up facts. Six puzzling behaviours, from hallucination to overconfidence, follow directly. Addressing the obvious objection that modern AI searches the Internet, it distinguishes model from product: retrieval changes what the model predicts over, not how it works, like an open-book exam. AI literacy research confirms such understanding must precede educational deployment.

KEYWORDS

Artificial Intelligence, Prediction, Stochastic Parrots

Most people carry one of two stories about artificial intelligence in their heads, and both are wrong.

The first story says AI is a digital brain: a thinking entity that understands our questions, knows the answers, and consults some vast internal encyclopaedia before replying. The second story, told with a dismissive wave, says it is "just autocomplete," a glorified copy-paste machine. The first story produces awe, the second contempt, and neither produces what educators actually need, which is the ability to predict what the tool will do before it does it.

After three years of running AI workshops for Malaysian teachers, I have become convinced that everything hinges on one idea, small enough to fit in a sentence and powerful enough that a person who absorbs it can derive nearly every strange behaviour of these systems on their own.

Here it is: the AI model predicts; *it does not look up*.

A system like ChatGPT was trained on one task, and one task only: given a stretch of text, predict what plausibly comes next, refined across billions of examples from the Internet. That is the whole job. Researchers at Princeton have argued that the only way to genuinely understand these systems is to start from exactly this fact, analysing them through the problem they were trained to solve, next-word prediction over Internet text, the way a biologist understands an

animal through the environment it adapted to (McCoy et al. 2024). Inside the model there is no fact-checking department, no database of verified truths being consulted. There is prediction, at a scale so vast that prediction begins to look like understanding. The researchers who famously described these systems as "stochastic parrots" made the point sharply: a language model stitches together sequences of linguistic forms according to probability, without grounding in meaning or truth, while we humans, wired to find intent in fluent language, cannot help reading comprehension into it (Bender et al. 2021).

Hold that single idea, and watch how much it explains. Why does AI fabricate references and invent court cases? Because plausible and true are different properties, and the machine optimises the first; a fake journal article with realistic authors and a well-formed DOI is a perfectly plausible continuation of "list five references." Why is it persuasive even when wrong? Fluency is plausibility, and plausibility is the thing it was built to produce. Why does it reflect bias? Because it predicts the patterns in human text, and the patterns include ours. Why does prompting matter? Because you are not querying a database; you are setting up a context whose most probable continuation you will receive. Why can it apparently reason? Because vast amounts of its training text are written human reasoning, and predicting reasoning-shaped text step by step often reproduces valid reasoning. And why does it almost never say "I don't know"? Because on the Internet it learned from, a confident question is rarely followed by an admission of ignorance.

Now, I can hear the objection, because I hear it in every workshop, usually within the first ten minutes.

> "But Dr., my AI does look things up. It searches the Internet. It gives me links."

The person raising it is not wrong about what they are seeing, and this is precisely where most explanations of AI fall apart, either by ignoring the objection or by surrendering to it. So let us take it seriously, because resolving it is the second half of understanding the tool.

The resolution is a distinction between the model and the product. The model is the prediction engine described above. The product, the app on your phone, is that engine surrounded by plumbing. When today's AI "searches the Internet," here is what actually happens:

- The system runs a conventional web search,
- Fetches the pages,
- Pastes their text into the model's working context.
- Then the model does the only thing it has ever done: predict the most plausible continuation, now of the retrieved text sitting in front of it.

Engineers call this *retrieval-augmented generation*, an architecture designed precisely because the bare model has no reliable access to facts (Lewis et al. 2020). The search happens outside the model. The model then writes a plausible continuation of the search results.

The atom therefore survives, refined: retrieval changes what the model predicts over, not how it works.

Every teacher already knows this situation intimately. It is the difference between a closed-book and an open-book exam, taken by the same student. The open book helps enormously; answers grounded in real text in front of you are far more likely to be accurate, which is why search-connected AI hallucinates much less. But the book does not install understanding, and anyone who has marked open-book exams knows the student who confidently misquotes the very page lying open on the desk. AI with search behaves exactly like that student, and for the same structural reason. It will sometimes cite a real webpage for a claim the page never makes, because that citation was a plausible continuation. It will smooth over contradictions between two sources rather than flag them, because smooth text is more probable than awkward text. And it inherits whatever the search dragged in: feed it a content-farm article, and it will fluently

continue content-farm nonsense. Search gives the guesser better material. It does not install a mechanism for truth.

If this all sounds too tidy, the prediction story makes testable predictions of its own, and they hold. The Princeton team reasoned that a true prediction machine should perform better whenever the correct answer happens to be common text, even on tasks where commonness is irrelevant. Asked to count letters in a list, a leading model scored 97 percent when the correct answer was 30, a number that appears constantly in written text, but only 17 percent when the answer was 29, a rarer number. Asked to decode a simple cipher, it succeeded 51 percent of the time when the hidden message was a common sentence and 13 percent when it was unusual (McCoy et al. 2024). Counting is counting; difficulty should not depend on the popularity of the answer. For a prediction machine, it does. No digital-brain story explains that result, and no bookshelf of retrieved documents changes it.

I should be honest about where the tidy story ends. Whether prediction at this scale amounts to some genuine form of understanding is a live scientific debate, with serious researchers on both sides and our old concepts straining to describe something we have never built before (Mitchell & Krakauer 2023). But educators do not need to settle the philosophy to use the tool well. We need to know what the system reliably does, predict over whatever is in front of it, and what it structurally lacks, a mechanism for truth. That is enough to work with.

And it is exactly where AI use in education must begin. The research on AI literacy is consistent on the ordering: the foundational layer is knowing and understanding what AI is and how it works, and only on top of that do we build using, evaluating, and creating with it (Ng et al. 2021). We routinely get this backwards, deploying the tools first and scheduling the "critical AI" workshop later, if ever. The major review of large language models in education warns from the other direction: the opportunities are real, but they materialise only when teachers and learners bring competence and critical awareness to the tool, and the risks compound when they do not (Kasneci et al. 2023). A teacher who believes the machine looks things up, even one reassured by the presence of search links, will treat its output as an encyclopaedia entry. A teacher who knows it predicts, with or without the open book, will treat the output as a capable draft from a fluent, well-read, occasionally confabulating assistant, and will check the quotes against the book itself. Same tool. Entirely different classroom.

In my workshops I compress all of this onto one slide: It does not search for answers; it guesses the most plausible continuation, and Internet search merely supplies material for the guess. I have watched that slide change how a room full of teachers reads every AI output for the rest of the day, links and all.

We do not let students light a Bunsen burner without teaching them what fire is. We should not hand them a plausibility engine, however well-connected to the Internet, while letting them believe it is an oracle. The machine predicts. The plumbing retrieves. Neither one knows. Teach that first, and much of the fog around AI in our classrooms lifts on its own.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 9459–9474.

McCoy, R. T., Yao, S., Friedman, D., Hardy, M. D., & Griffiths, T. L. (2024). Embers of autoregression show how large language models are shaped by the problem they are trained to solve. Proceedings of the National Academy of Sciences, 121(41). https://doi.org/10.1073/pnas.2322420121

Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. Proceedings of the National Academy of Sciences, 120(13). https://doi.org/10.1073/pnas.2215907120

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041