

The human knowledge loophole in the 'bitter lesson' for LLMs

ICLR 2026 (blog post track)

Anna Rogers

The 'bitter lesson' of R. Sutton

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin.

examples: chess, Go, speech recognition, computer vision

Sutton R. (2019) [The bitter lesson.](#)

(at least) two readings:

- ‘hard’: scale is all you need
- ‘soft’: approaches relying on human knowledge have historically been beaten by those that instead rely on computation

? whose knowledge is it?

Rogers A. (2026) [The human knowledge loophole in the 'bitter lesson' for LLMs](#)

'Bitter lesson' in NLP: pre-deep-learning methods

"Every time I fire a linguist, the performance of the speech recognizer goes up." - F. Jelinek

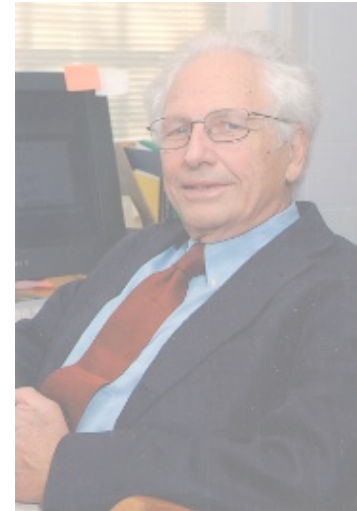
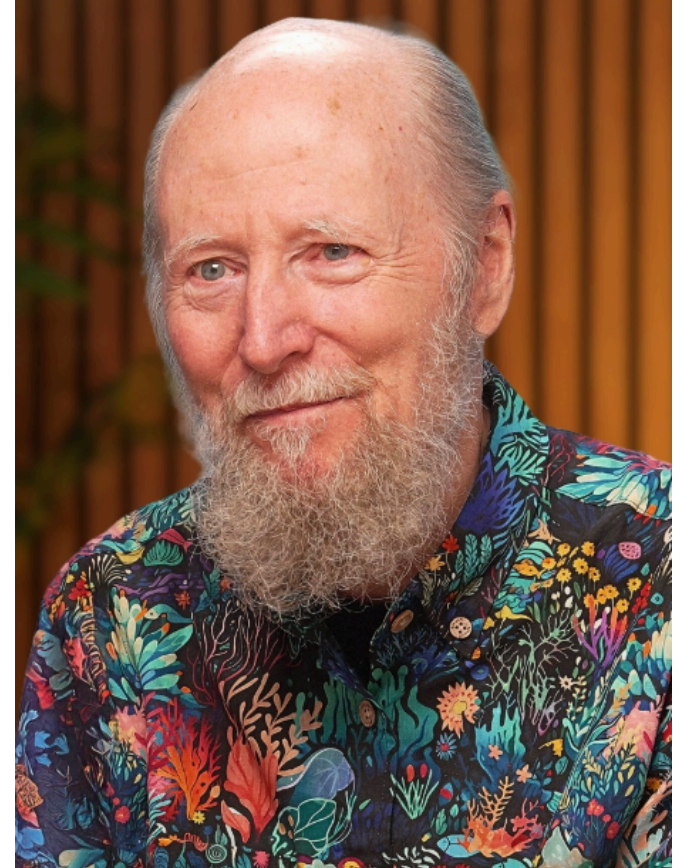


image: IT history society <https://www.ithistory.org/honor-roll/frederick-jelinek>

Is the success of LLMs due to non-reliance on human knowledge?

It's an interesting question whether LLMs are a case of the bitter lesson. They are clearly a way of using massive computation... But they're also a way of putting in lots of human knowledge.



R. Sutton (Sep. 26 2025) Dwarkesh Podcast

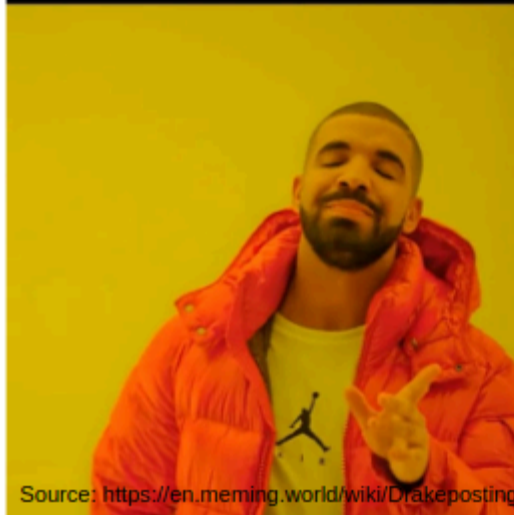
If more compute is the solution, why do we have all this?

- Why does **scaling** require the scaling of both compute and data?
- Why do models **consistently perform worse out-of-distribution**?
- Why do we **optimize data mixes for benchmarks**?
- ...

GOFAI vs deep learning



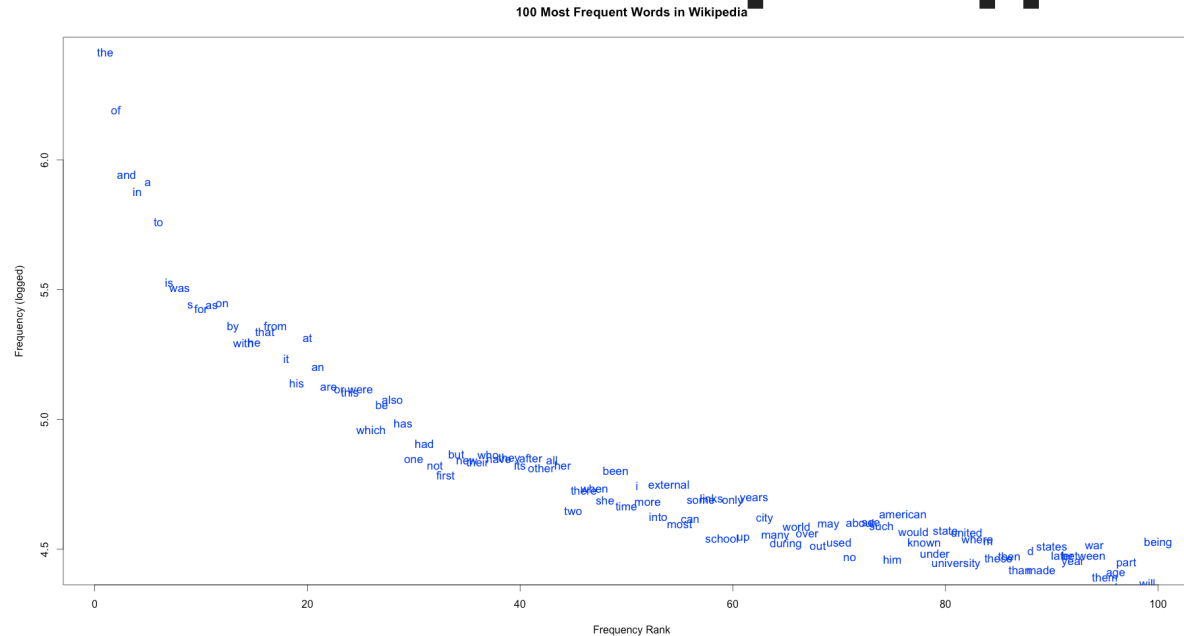
a hand-written
rule
for each
phenomenon



hundreds of
hand-written
examples
for each
phenomenon

Source: <https://en.meming.world/wiki/Drakeposting>

Why do LLMs work better than prior approaches?



- 👍 in this distribution we easily cover frequent patterns
- 👎 not all frequent patterns are desirable
- 👎 patterns in the long tail neither covered nor known

Image credit: <http://wugology.com/zipfs-law/>

What about synthetic data?

it can improve representation of *form*, not *content*!

- 👍 we can generate more examples of summaries or QA pairs to 'teach' these task formats
- 👎 we can't make up the facts in the summaries or the linguistic patterns, it has to *continually* come from humans



Can a true 'bitter lesson' solution exist for language?

- dependence on human knowledge is unavoidable when human knowledge is a big part of what is being modeled
- the original Sutton's analogy with games, where information is free, does not serve us well in NLP
- we need learning methods that would support faithful and computationally cheap data attribution!

Rogers A. (2026) [The human knowledge loophole in the 'bitter lesson' for LLMs](#)

Thank you!

 arog@itu.dk

Comments and collaborations welcome!

[@annarogers.bsky.social](https://bsky.social/@annarogers)

 <https://linkedin.com/in/annargrs/>