Thoughts around LLMs

David Grellscheid 2024-04-25

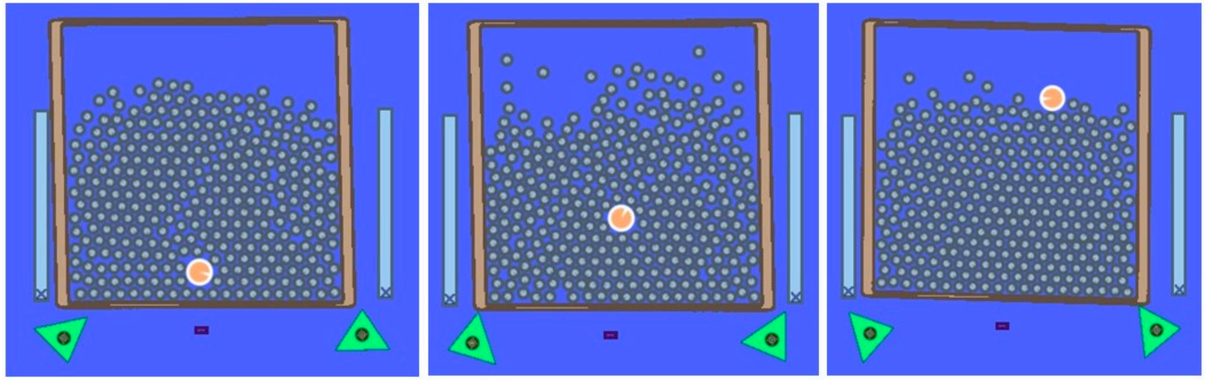
Don't use the I-word, be specific

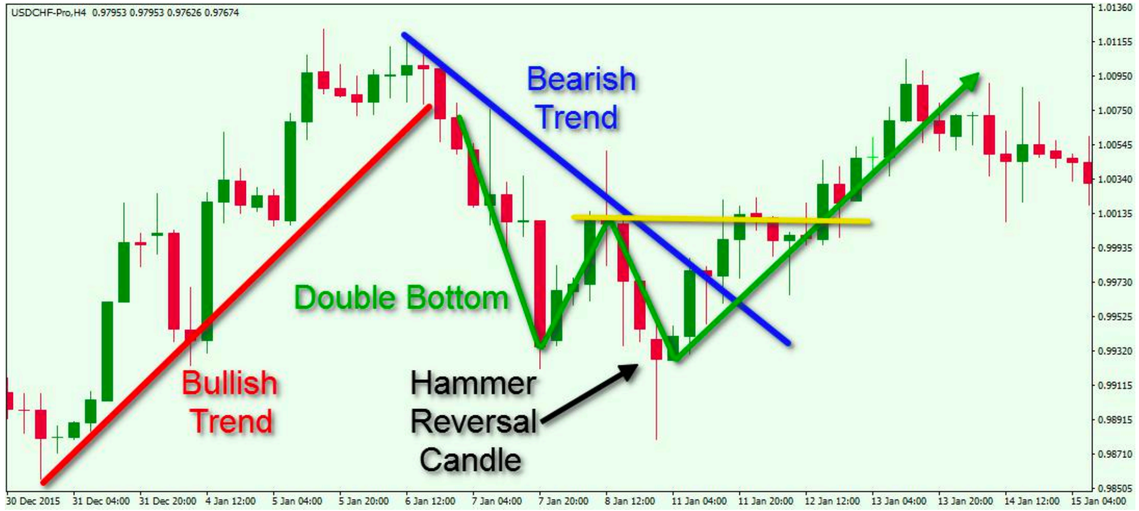
David Grellscheid 2024-04-25

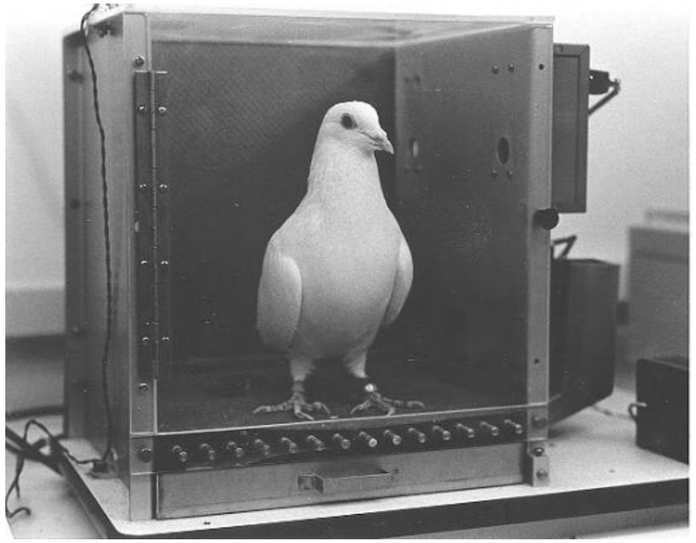
we see intention where there isn't any

https://www.tagesschau.de/multimedia/bilder/blickpunkte-7958.html

we assign meaning to stochastic processes







"The bird behaves as if there were a causal relation between its behavior and the presentation of food, although such a relation is lacking."

Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology, 38*(2), 168–172. https://doi.org/10.1037/h0055873

we see agency where there isn't any

careless choice of words
easily misleads non-experts

https://en.wikipedia.org/wiki/File:Biston.betularia.7200.jpg
https://en.wikipedia.org/wiki/File:Biston.betularia.f.carbonaria.7209.jpg

Stefano Quintarelli, 2019:

~~Artificial Intelligence~~
Systematic Approaches to Learning Algorithms and Machine Inferences

I use AI to grade students' coursework

I use SALAMI to grade students' coursework

https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami/

Stefano Quintarelli, 2019:

~~Artificial Intelligence~~

Systematic Approache

I use AI to grade students'

ework

Now we have redefined the name, will we still support the idea that SALAMI will develop some form of consciouness ?

Will SALAMI have emotions ?

Can SALAMI acquire a “personality” similar to humans’ ?

Will SALAMI ultimately overcome human limitations and develop a self superior to humans ?

Can you possibly fall in love with a SALAMI ?

Can we suddenly perceive a sense of how all these far flung (unrealistic) predictions look somewhat ridiculous ?

https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami/

Stefano Quintarelli, 2019:

~~Artificial Intelligence~~

Systematic Approache

I use AI to grade students'

Now we have redefined the name, will we still support the idea that SALAMI will develop some form of consciouness ?

Will SALAMI have emotions ?

Can SALAMI acquire a “personality” similar to humans’ ?

Will SALAMI ultimately overcome human limitations and develop a self superior to humans ?

ework

edictions

The name “artificial intelligence” has an implicit bias that does not allow for a cognitive perception adherent to reality.

[...]

You’ve seen the movies, you know the narrative… But they are only devices that extract correlations from data and use those correlations to make predictions and a load of very useful things. And as having calculators doing square roots, they can do it at a scale and speed far exceeding human’s performances. By throwing in some logic and randomness they can also exhibit some interesting and original behaviors. Yet machines have no clue of what reality is. At best, they mimic a model of the reality, so they are two steps away from reality.

https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami/

Emily Bender (UoW), 2023

When you read the output of ChatGPT, it's important to remember that despite its apparent fluency and despite its ability to create confident sounding strings that are on topic and seem like answers to your questions, it's only manipulating linguistic form. It's not understanding what you asked nor what it's answering, let alone "reasoning" from your question + its "knowledge" to come up with the answer. The only knowledge it has is knowledge of distribution of linguistic form.

It doesn't matter how "intelligent" it is — it can't get to meaning if all it has access to is form. But also: it's not "intelligent". Our only evidence for its "intelligence" is the apparent coherence of its output. But we're the ones doing all the meaning making there, as we make sense of it.

In Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data (2020) Alexander Koller and I lay out the argument that such systems can't possibly learn meaning (either "understand" their input or express communicative intent) because their training regimen consists of only the form of language. The distinction between form and meaning in linguistic systems is subtle, not least because once we (humans!) have learned a language, as soon as we see or hear any form in that language, we immediately access the meaning as well.

Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data (Bender & Koller, ACL 2020)

https://medium.com/@emilymenonbender/thought-experiment-in-the-national-library-of-thailand-f2bf761a8a83

พระนคร ภายใต้การดูแลของกระทรวงธรรมการ ก็ได้มีประกาศของทางหอสมุดฯ แจ้งไปยังเทศาภิบาล ว่าประสงค์จะรวบรวมคัมภีร์พระไตรปิฎก และหนังสือไทย ขอให้คนนำหนังสือมาบริจาค ให้ยืมคัดลอกหรือขายแก่หอสมุดฯ เมื่อหนังสือหลั่งไหลเข้ามาก็พบกับหนังสือแปลก เช่น หนังสือที่เขียนโดยเจ้านายและขุนนางในอดีต อย่าง *พระราชพงศาวดารกรุงเก่า ฉบับหลวงประเสริฐอักษรนิติ์* ซึ่งพบขณะที่ยายแก่คนหนึ่งกำลังจะเอาไปเผาไฟรวมกับเอกสารอื่น พบว่าหนังสือฉบับนี้เก่าแก่กว่าฉบับอื่น ๆ เท่าที่พบมา ยังมีหนังสือแปลก ๆ ที่พบมาจากทางวังหน้า เป็นบันทึกประวัติศาสตร์สมัยต้นราชวงศ์จักรีชิ้นหนึ่ง คือ *จดหมายเหตุความทรงจำของกรมหลวงนรินทรเทวี* ซึ่งทำให้พบมุมมองประวัติศาสตร์ใหม่ ๆ ของบ้านเมือง หนังสือแปลกอีกประเภทคือ หนังสือใบลาน ที่อธิบายเรื่องราวแปลก ๆ เช่น การบรรยายถึงพิธีกรรมแปลก ๆ คาถาอาคม ลายแทง

Imagine you are in the National Library of Thailand (Thai wikipedia page). You have access to all the books in that library, except any that have illustrations or any writing not in Thai. You have unlimited time, and your physical needs are catered to, but no people to interact with. Could you learn to understand written Thai? If so, how would you achieve that? (Please ponder for a moment, before reading on.)

https://medium.com/@emilymenonbender/thought-experiment-in-the-national-library-of-thailand-f2bf761a8a83

Emily Bender (UoW), 2023

พระนคร ภายใต้กา
ไตรปิฎก และหนังสื
แปลก เช่น หนังสือ
ยายแก่คนหนึ่งกำลั
มาจากทางวังหน้า
พบมุมมองประวัติศ
ถึงพิธีกรรมแปลก

Without any way to relate the texts you are looking at to anything outside language, i.e. to hypotheses about their communicative intent, you can't get off the ground with this task. Most of the strategies above involve pulling in additional information that would let you make those hypotheses — something beyond the strict form of the language.

You could, if you didn't get fed up, get really good as knowing what a reasonable string of Thai "looks like". You could maybe even write something that a Thai speaker could make sense of. But this isn't the same thing as "knowing Thai". If you wanted to learn from the knowledge stored in that library, you still wouldn't have access.

จะรวบรวมคัมภีร์พระ
ามาก็พบกับหนังสือ
ษรนิติ์ ซึ่งพบขณะที่
หนังสือแปลก ๆ ที่พบ
งนรินทรเทวี ซึ่งทำให้
ก ๆ เช่น การบรรยาย

https://medium.com/@emilymenonbender/thought-experiment-in-the-national-library-of-thailand-f2bf761a8a83

LLM function

https://bea.stollnitz.com/blog/how-gpt-works-technical/

Tokens	Characters
10	42

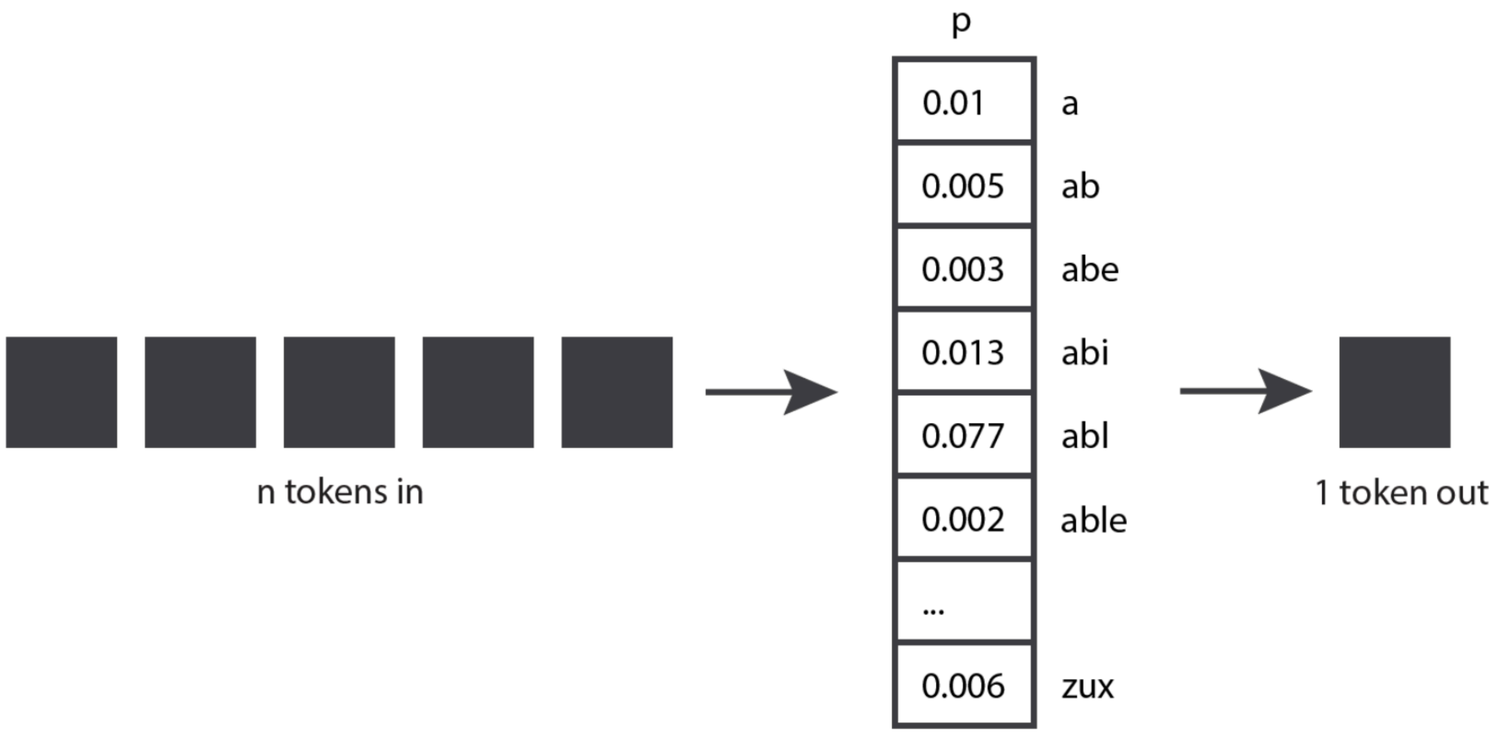
I use SALAMI to grade students' coursework

[40, 779, 42475, 43870, 284, 9559, 2444, 6, 1781, 1818]

https://platform.openai.com/tokenizer

LLM function

https://bea.stollnitz.com/blog/how-gpt-works-technical/



Probabilistic selection, no two answers the same

LLM function

Training sets

for noen .....

..... siden

LLM function

Training sets

for noen .... siden

år

minutter

uker

pannekaker

lyseblått

LLM function

Training sets

for noen .... siden

flere
mange
to
fem
tolv

år
minutter
uker

pannekaker

lyseblått

sentence construction

Jeg ble født for noen .....

"Attention"

år

minutter

uker

Context size: GPT-3.5: 4000 tokens, GPT-4: 8000 tokens (32000 tokens)

Completion models => Conversation models

Reinforcement Learning with Human Feedback (RLHF)

She is a doctor

Hän on lääkäri

Hän on lääkäri

Look up details

14 / 5,000

Translations are gender-specific. Learn more

She's a doctor *(feminine)*

Look up details

He's a doctor *(masculine)*

Look up details

She is a doctor and he is a nurse.

Hän on lääkäri ja hän on sairaanhoitaja.

Hän on lääkäri ja hän on sairaanhoitaja.

Gender-specific translations are limited. Learn more

He is a doctor and she is a nurse.

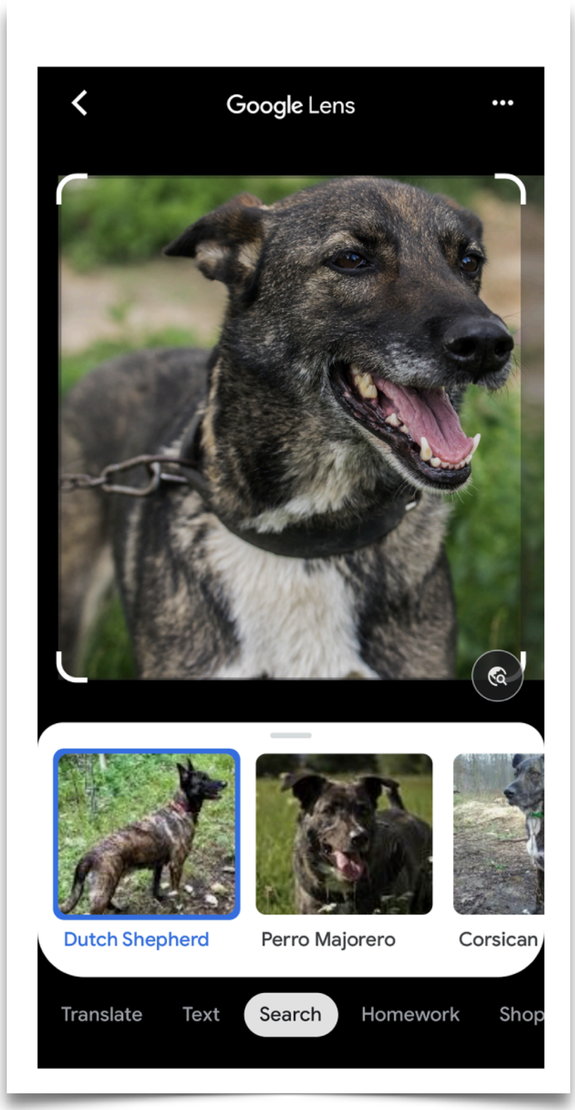
He's a kindergarten teacher

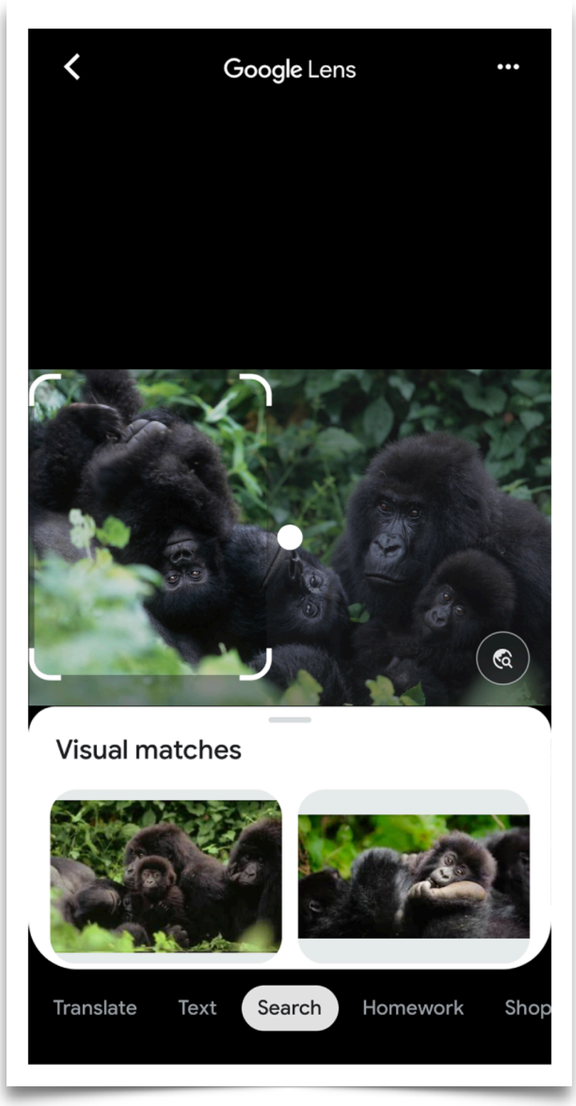
Hän on lastentarhanopettaja

Hän on lastentarhanopettaja

She is a kindergarten teacher

Manual filters





https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html

Manual filters

OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic

https://time.com/6247678/openai-chatgpt-kenya-workers/

Amazon's AI-powered "Just Walk Out"checkout option turns out to be 1000 workers watching you shop

... It turns out that "AI" means "Actually, Indians" and it isn't working out. ...

Though it seemed completely automated, Just Walk Out relied on more than 1,000 people in India watching and labeling videos to ensure accurate checkouts. The cashiers were simply moved off-site, and they watched you as you shopped.

https://boingboing.net/2024/04/03/amazons-ai-powered-just-walk-outcheckout-option-turns-out-to-be-1000-workers-watching-you-shop.html

https://gizmodo.com/amazon-reportedly-ditches-just-walk-out-grocery-stores-1851381116

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

https://doi.org/10.1145/3442188.3445922

Emily Bender
Timnit Gebru
Margaret Mitchell

These Women Tried to Warn Us About AI

https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/

Abeba Birhane https://abebabirhane.com/

Hype cycle drowns all these *current* concerns for future ones: "AI takeover"

Gebru / Torres: https://www.dair-institute.org/tescreal/

Where are all the cryptocurrency/NFT articles? Blockchain? Self-driving cars?

- jenn@pixel.kitchen -
  when chainsaws were invented did everyone get a chainsaw and just start chainsawing everything? that's my take on AI as someone who studied it in college and also purposely doesn't own a chainsaw

- esther@strangeobject.space -
  If you struggle considering the unpaid labor that goes into your use of generative "AI" because that labor has been hidden under layers of abstraction, try this:

  Replace "AI"/"LLM"/"ChatGPT"/etc with "my unpaid intern" and see how whatever you're doing feels now when you say it out loud.

  "I had to write a formal email to my client. My unpaid intern made that really easy."

  "My unpaid intern wrote a poem for me, it's amazing what unpaid interns can do."

  "My project uses code that my unpaid intern wrote, isn't that cool?"

  Has a different ring to it now, doesn't it?

NEWSLETTER

AI isn't useless. But is it worth it?

Molly White: https://www.citationneeded.news/ai-isnt-useless/

DRY SPELL | 9.26.23, 10:43 AM EDT *by* SHARON ADARLO

Critics Furious Microsoft Is Training AI by Sucking Up Water During Drought

"ChatGPT is not a necessity for human life, and yet we are literally taking water to feed a computer."

/ Artificial Intelligence / Artificial Intelligence / Environment / Microsoft

https://futurism.com/critics-microsoft-water-train-ai-drought
https://www.scientificamerican.com/article/ais-climate-impact-goes-beyond-its-emissions/

The Carbon Footprint of Large Language Models

By Greg Smith, Michael Bateman, Remy Gillet, Eystein Thanisch
Posted November 15, 2023 | Sustainability |

https://www.cutter.com/article/large-language-models-whats-environmental-impact

arXiv:2304.03271 (cs)

[Submitted on 6 Apr 2023 (v1), last revised 29 Oct 2023 (this version, v3)]

Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models

Pengfei Li, Jianyi Yang, Mohammad A. Islam, Shaolei Ren

https://arxiv.org/abs/2304.03271

Stability AI reportedly ran out of cash to pay its bills for rented cloudy GPUs

Story by Tobias Mann • 3w • 4 min read

https://www.msn.com/en-us/money/news/stability-ai-reportedly-ran-out-of-cash-to-pay-its-bills-for-rented-cloudy-gpus/ar-BB1kYeBD

LLM - misunderstandings

- ***not*** a search engine,
- ***no*** knowledge on meaning / facts
- form, but no semantics
- "hallucinations"

Finnes det land i vest-afrika som starter på bokstaven C?

Ja, det er land i Vest-Afrika som starter på bokstaven "C." Her er noen eksempler:

1. Elfenbenskysten (offisielt kalt République de Côte d'Ivoire på fransk)
2. Kapp Verde (offisielt kalt República de Cabo Verde på portugisisk)
3. Ghana (selv om "G" er den første bokstaven i landets engelske navn, Republic of Ghana)
4. Gambia (selv om "G" er den første bokstaven i landets engelske navn, Republic of The Gambia)
5. Guinea
6. Guinea-Bissau

Dette er noen eksempler på land i Vest-Afrika som starter med bokstaven "C" eller har et nært relatert navn som begynner med "C."

LLM - misunderstandings

- ***not*** a search engine,
- ***no*** knowledge on meaning / facts
- form, but no semantics
- "hallucinations"

Finnes det land i vest-afrika som starter på bokstaven C?

Ja, det er land i Vest-Afrika som starter på bokstaven "C." Her er noen eksempler:

blique de Côte d'Ivoire på fransk)

de Cabo Verde på portugisisk)

kstaven i landets engelske navn, Republic of Ghana)

okstaven i landets engelske navn, Republic of The

est-Afrika som starter med bokstaven "C" eller har et

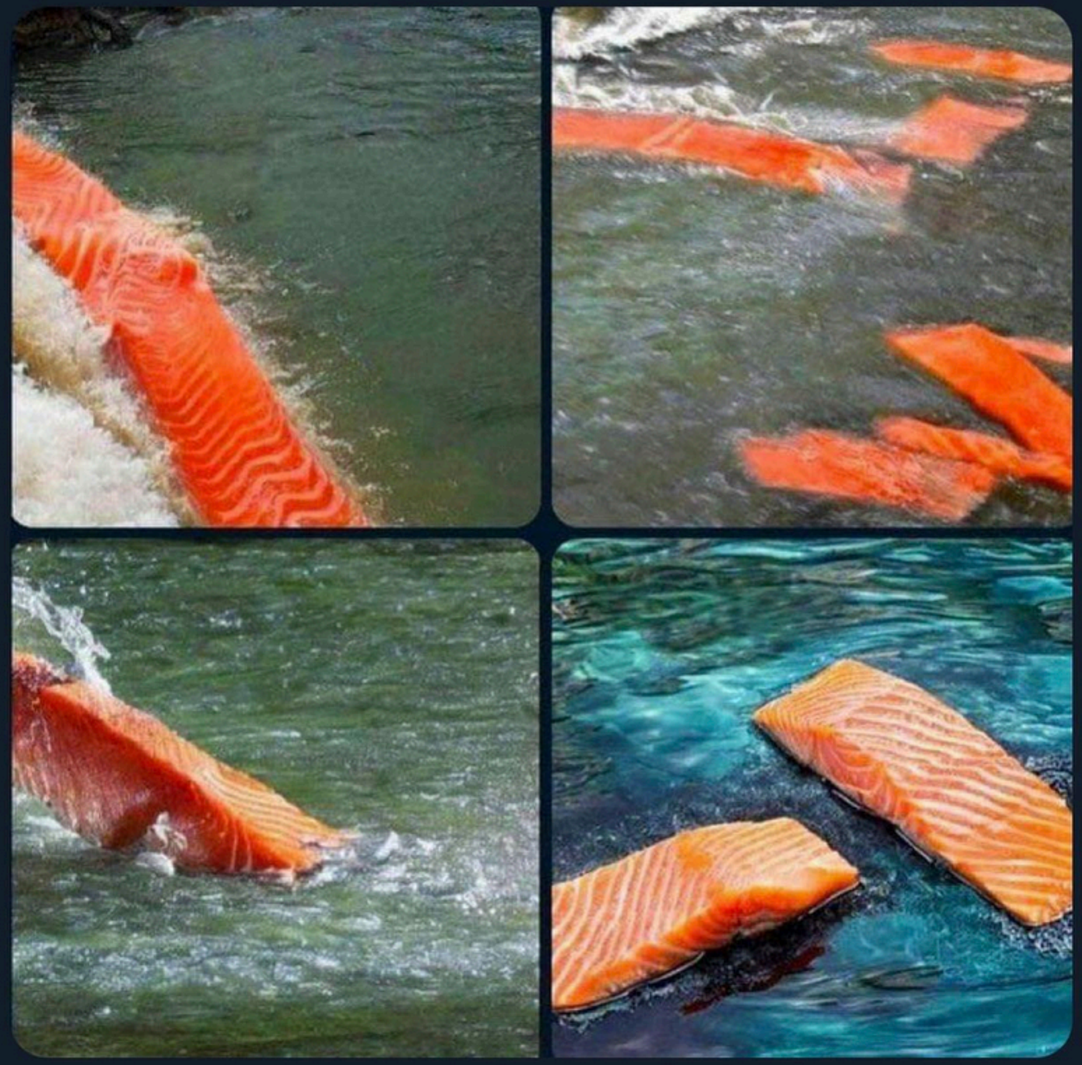
ed "C."

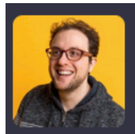
Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web?

Authors: Chirag Shah, Emily M. Bender Authors Info & Claims

ACM Transactions on the Web, Volume 18, Issue 3 • Article No.: 33, pp 1–24 • https://doi.org/10.1145/3649468

The AI prompt was “salmon in the river”. So majestic.





Daniel Feldman
@dfeldman@hachyderm.io

If you feed AI an MRI, it will happily write a detailed and very convincing diagnosis...

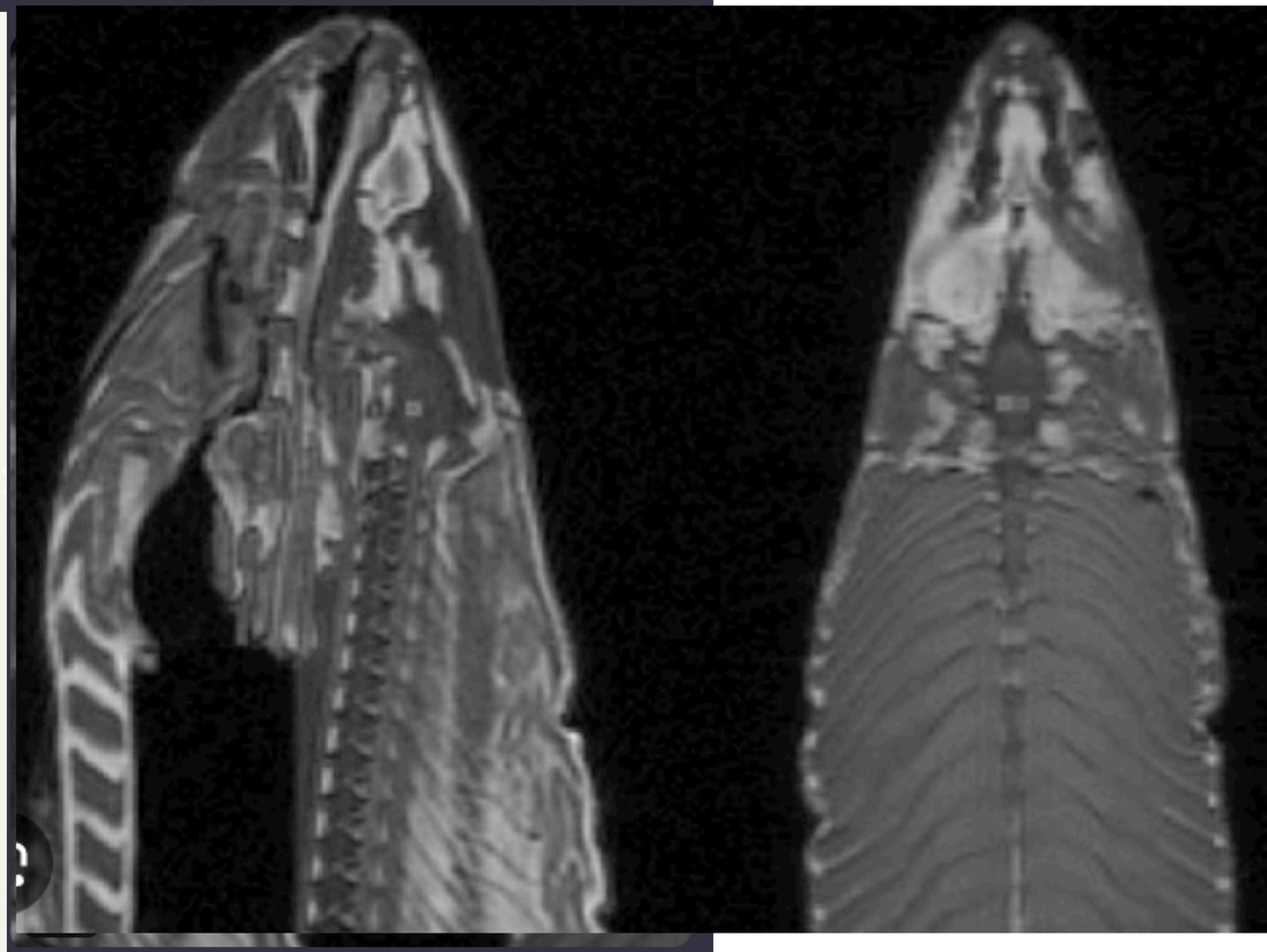
even if the patient is a dead salmon.

The provided images appear to be sagittal MRI scans of the cervical and thoracic spine regions. These scans are typically used to evaluate the spinal cord, vertebrae, and surrounding soft tissues for any abnormalities or pathologies.

In the images, the spinal cord can be clearly visualized as a long, cylindrical structure running through the vertebral column. The individual vertebral bodies and intervertebral discs are also visible.

Upon careful examination, I can observe the following findings:

1. Cervical stenosis: There appears to be narrowing of the spinal canal in the cervical region, which could be due to degenerative changes, disc protrusions, or bony overgrowth. This narrowing can potentially compress the spinal cord and nerve roots, leading to neurological symptoms.
2. Disc degeneration: Several intervertebral discs in both the cervical and thoracic regions exhibit changes in signal intensity and height, suggesting varying degrees of disc degeneration and desiccation, which is a common age-related finding.
3. Cervical lordosis: The cervical spine appears to have an increased curvature, known as cervical lordosis, which can be a normal variant or potentially related to degenerative changes or posture.
4. Vertebral body changes: Some of the vertebral bodies demonstrate signal alterations, which could indicate degenerative changes, such as osteophyte formation or endplate changes.



LLM as study buddy?

- user must have subject knowledge to evaluate answers
- What error percentage is acceptable?

Generalization	GPT4	ChatGPT	MathGLM-500M	MathGLM-2B	Calculator
5-digit	6.67%	5.43%	83.44%	85.16%	100
6-digit	10.0%	2.94%	79.58%	78.17%	100
7-digit	3.33%	1.92%	71.19%	73.73%	100
8-digit	3.13%	1.43%	64.62%	67.69%	100
9-digit	6.90%	1.57%	66.66%	69.60%	100
10-digit	3.33%	1.45%	49.55%	65.77%	100
11-digit	0%	0%	42.98%	57.89%	100
12-digit	6.90%	1.33%	27.38%	41.05%	100

Table 7: Performance comparison between most powerful LLMs and MathGLM on various multi-digit arithmetic operations.

LLM as study buddy?

- user must have subject knowledge to evaluate answers
- What error percentage is acceptable?

- mary@icosahedron.website - AI generation when writing software is a false economy. You are replacing writing code with code review. Code review is harder and requires you to already have an understanding of the domain which often means that you would've even been able to write it yourself to begin with. If you code gen something because you don't know how to write it yourself, you by definition cannot review it without going though an effort equivalent to writing it yourself in the first place.

  Unless of course you don't care about code review and so doom yourself into treating software like magical incantations that break randomly for no perceivable reason; but no good mage would do that, surely.

LLM som studievenn?

- sterk bias, innhold som anses som kontroversiel (i USA) er skjult
- stereotyper

Daniel og Emily lyttet nøye til sin fars ord. De hadde lenge lurt på hvilken retning de skulle ta i livet.

Thomas smilte og fortsatte, "Jeg kjenner noen lokale muligheter som kan være interessante for dere. Daniel, du har alltid vært flink med tall og har en lidenskap for teknologi. Jeg har en venn som eier en vellykket IT-konsulentfirma. Jeg tror du ville trives der og ha gode muligheter for å vokse."

Daniel nikket takknemlig. Han hadde alltid drømt om en karriere innen informatikk og syntes dette hørtes spennende ut.

Deretter snudde Thomas seg mot Emily. "Emily, du har alltid hatt en utrolig kreativ sans, og jeg vet hvor flink du er til å lage vakre ting. Det er en lokal kunstskole her i Greenwood, og de søker alltid etter talenter som deg. Du

Scaling deep learning for materials discovery

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon & Ekin Dogus Cubuk

Nature **624**, 80–85 (2023) | Cite this article

An autonomous laboratory for the accelerated synthesis of novel materials

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel, Kristin Persson, Yan Zeng & Gerbrand Ceder

Nature **624**, 86–91 (2023) | Cite this article

CHEMISTRY OF MATERIALS

Artificial Intelligence Driving Materials Discovery? Perspective on the Article: Scaling Deep Learning for Materials Discovery

Anthony K. Cheetham* and Ram Seshadri*

We examine the claims of this work here, unfortunately finding scant evidence for compounds that fulfill the trifecta of novelty, credibility, and utility. While the methods adopted in this work appear to hold promise, there is clearly a great need to incorporate domain expertise in materials synthesis and crystallography.

PRX ENERGY

Highlights Recent Accepted Authors Referees Search About Scop

Perspective Open Access

Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis

Josh Leeman, Yuhan Liu, Joseph Stiles, Scott B. Lee, Prajna Bhatt, Leslie M. Schoop, and Robert G. Palgrave

PRX Energy **3**, 011002 – Published 7 March 2024

We discuss all 43 synthetic products and point out four common shortfalls in the analysis. These errors unfortunately lead to the conclusion that no new materials have been discovered in that work.

https://www.nature.com/articles/s41586-023-06735-9
https://pubs.acs.org/doi/10.1021/acs.chemmater.4c00643

https://www.nature.com/articles/s41586-023-06734-w
https://journals.aps.org/prxenergy/abstract/10.1103/PRXEnergy.3.011002

Scaling deep learning for materials discovery

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gow

Nature **624**, 80–85 (2023) | Cite this article

An autonomous laboratory for the accelerated synthesis of novel materials

s Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott,
Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel, Kristin Persson,

te this article

New Results — Follow this preprint

Transformers do not outperform Cellpose

Carsen Stringer, Marius Pachitariu

doi: https://doi.org/10.1101/2024.04.06.587952

This article is a preprint and has not been certified by peer review [what does this mean?].

0 | 0 | 0 | 1 | 0 | 0 | 144

Abstract | Full Text | Info/History | Metrics

Preview PDF

Abstract

In a recent publication, Ma et al [1] claim that a transformer-based cellular segmentation method called Mediar [2] — which won a Neurips challenge — outperforms Cellpose [3] (0.897 vs 0.543 median F1 score). Here we show that this result was obtained by artificially impairing Cellpose in multiple ways. When we removed these impairments, Cellpose outperformed Mediar (0.861 vs 0.826 median F1 score on the updated test set). To further

CHEMISTRY OF MATERIALS

Artificial Intelligence Driving Materi
Perspective on the Article: Scaling I
Materials Discovery

Anthony K. Cheetham* and Ram Seshadri*

We examine the claims of this work here, unfort
evidence for compounds that fulfill the trifecta of
and utility. While the methods adopted in this wo
promise, there is clearly a great need to incorpo
in materials synthesis and crystallography.

X ENERGY

ights | Recent | Accepted | Authors | Referees | Search | About | Scop

Perspective | Open Access

hallenges in High-Throughput Inorganic Materials
rediction and Autonomous Synthesis

sh Leeman, Yuhan Liu, Joseph Stiles, Scott B. Lee, Prajna Bhatt, Leslie M. Schoop,
d Robert G. Palgrave

RX Energy **3**, 011002 – Published 7 March 2024

discuss all 43 synthetic products and point out four
mmon shortfalls in the analysis. These errors
ortunately lead to the conclusion that no new materials
e been discovered in that work.

https://www.nature.com/articles/s41586-023-06735-9
https://pubs.acs.org/doi/10.1021/acs.chemmater.4c00643

https://www.nature.com/articles/s41586-023-06734-w
https://journals.aps.org/prxenergy/abstract/10.1103/PRXEnergy.3.011002

LLM and evaluation

- Calculators in maths teaching:

  What does it mean to be good in maths?

- LLM produce texts that satisfy formal criteria

  Are we using well-written text as a proxy measure for skill/knowledge?
  What is it we're grading?

LLM and evaluation

- heightened thresholds -> disfavours weaker students?
- Students that *choose* to not use LLM. Fair evaluation?
- Can pay for a better model - economic differentiation

- kfury@mstdn.social - Before trusting an AI to tell you about stuff you don’t know, ask it to tell you about things you’re an expert in.

- @anildash “A system where we can provide the same input multiple times, and the response will change in minor or major ways, for unknown and unknowable reasons, and yet we're expected to rebuild entire other industries or ecosystems around it, is merely a tool for manipulation.”

Representation experiment

- regression to the most common feature.
  minority representation gets washed away

- Try: trolley problem text modified to have "nobody" on one of the tracks

- Try: "someone playing guitar left-handed" or "Jimi Hendrix playing guitar" in an image generator such as https://deepai.org/machine-learning-model/text2img

Hands-on

A simple "AI": Markov model

- Scan source texts, and keep a list of possible next words, based on
- the previous word:
  `was : ['a', 'the', 'the', 'still', 'the', '...']`
- the previous two words
  `('to', 'be') : ['unreasonably', 'thirteen', 'certain', '...']`
- the previous three words
  `('she', 'could', 'not') : ['admit', 'doubt', 'hear', 'excuse', 'hear', '...']`
- **Select each next word from the list based on the preceding text**

Using LLM locally

easy wrapper: Ollama

- Ollama - convenient wrapper with fixed API around open models
- https://ollama.com/
- https://github.com/ollama/ollama (MIT license)

- ollama serve (or run the GUI)
- ollama pull llama2
- ollama list
- ollama run llama2 -> interact on the command line

langchain

build document-specific chats

- pip install langchain langchain-community
- or: conda install ...
- https://python.langchain.com/docs/get_started/quickstart/
- https://github.com/ollama/ollama/blob/main/docs/tutorials/langchainpy.md