

Opportunities for the Future

ACL Tutorial, “A gentle introduction to deep nets, and opportunities for
the future”

Part B

May 22, 2022

Language Models (LMs)

History:

- Firth (1957)
 - *You shall know a word by the company it keeps*
- PMI (Church & Hanks, 1990)
- Word2Vec
- BERT

By construction, methods based on collocations (words that appear near one another more than chance):

- Strengths:
 - fluency
 - word-associations (psychology)

And Weaknesses:

- Truth:
 - logical form
 - temporal/spatial logic
 - possible worlds
- Meaning
- Purpose
 - Planning/Discourse Structure
- Common Sense Knowledge

General disclaimer

Language Models (LMs) are a very rapidly moving target.

Necessarily, some of our examples are from systems that are not current as of May 22, 2022. [This has risks.](#)

Some of the specific examples may work better, or work well, on LMs that exist today.

But there is no reason to believe the underlying problems are solved.

Semantics (Negation)

[“What BERT is not”](#) (Allyson Ettinger, TACL-2020)

- Given “A robin is a ____.” BERT fills in “bird” 
- Given “A robin is not a ____.” ditto 

Semantics (Negation)

“What BERT is not” (Allyson Ettinger, TACL-2020)

- Given “A robin is a ____.” BERT fills in “bird” 
- Given “A robin is not a ____.” ditto 

Collocations, Synonyms & Antonyms

- *You shall know a word by the company it keeps* – Firth (1957)
- By construction, embeddings (PMI, Word2Vec, BERT)
 - find words that appear near one another (more than chance)
 - including both synonyms as well as antonyms
 - (since corpora are full of comparisons & contrasts)
- In Word2Vec, *moral* is closer to *immoral* than to *good*.
 - (since *moral* and *immoral* appear in similar contexts more than near *good*)
 - Collocation ≠ Synonymy

Reliability / Truthfulness

To the extent that a use case places importance on the truth of the outputs provided, it is not a good fit for GPT-3 – Dale (JNLE, 2020)

[GPT-3: What's it good for?](#)

Journal of Natural Language Engineering

Reliability / Truthfulness

To the extent that a use case places importance on the truth of the outputs provided, it is not a good fit for GPT-3 – Dale (JNLE, 2020)

GPT-3: What's it good for?

Journal of Natural Language Engineering



Most read

This page lists the top ten most read articles for this journal based on the number of full text views and downloads recorded on Cambridge Core over the last 30 days. This list is updated on a daily basis.

- [GPT-3: What's it good for?](#)
Robert Dale
Published online by Cambridge University Press: 15 December 2020, pp. 113-118
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)
- [Word2Vec](#)
KENNETH WARD CHURCH
Published online by Cambridge University Press: 16 December 2016, pp. 155-162
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)
- [Emerging trends: Ethics, intimidation, and the Cold War](#)
Kenneth Ward Church, Valia Kordoni
Published online by Cambridge University Press: 31 May 2021, pp. 379-390
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)
- [Natural language processing in mental health applications using non-clinical texts](#)
RAFAEL A. CALVO, DAVID N. MILNE, M. SAZZAD HUSSAIN, HELEN CHRISTENSEN
Published online by Cambridge University Press: 30 January 2017, pp. 649-685
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)
- [The return of the chatbots](#)
ROBERT DALE
Published online by Cambridge University Press: 13 September 2016, pp. 811-817
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)
- [Emerging trends: A gentle introduction to fine-tuning](#)
Kenneth Ward Church, Zeyu Chen, Yanjun Ma
Published online by Cambridge University Press: 26 October 2021, pp. 763-778
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)
- [Emerging Trends: SOTA-Chasing](#)
Kenneth Ward Church, Valia Kordoni
Published online by Cambridge University Press: 08 February 2022, pp. 249-269
[Article](#) [Access](#) [Open access](#) [PDF](#) [HTML](#) [Export citation](#)
[View abstract](#)



Hallucinations / Possible Worlds

- LMs have a tendency to “hallucinate” when summarizing documents
- Output sounds plausible,
 - but may add embellishments to the input
- LMs tend to make up “alternative facts”
 - faster than they can be fact-checked
- This may well be their most dangerous failing
- People might believe some of these conspiracy theories

Reliability:

Experiment run by Ken Church

- **In 1885, Stanford University was founded** by Leland and Jane Stanford in memory of their only child, recently deceased.
- In 1887, The first successful gasoline powered automobile in the United States was manufactured in Cleveland, Ohio by Charles and Frank Duryea.
- In 1887, The American Federation of Labor was founded.
- In 1887, The Salvation Army was formally incorporated in London, England.
- In 1890, The American Red Cross was established by Clara Barton.

...

- In 1893, The first Miss America Pageant was held in New York City.
- In 1895, The National Association for the Advancement of Colored People(NAACP) was founded.
- In 1895, The National Council of Women of the United States was founded.
- In 1897, The Boston Symphony Orchestra gave its first performance.
- In 1897, The first winner of the Nobel Peace Prize was born.
- In 1899, The first automobile race was held at Narragansett Park in Cranston, Rhode Island.
- In 1900, The first automobile race was held at Narragansett Park in Cranston, Rhode Island.
- In 1900, The National Association for the Advancement of Colored People (NAACP) was founded.

Looks great!

But:

In 1893, The first Miss America Pageant was held in New York City.

In 1895, The National Association for the Advancement of Colored People (NAACP) was founded.

In 1895, The National Council of Women of the United States was founded.

In 1897, The Boston Symphony Orchestra gave its first performance.

In 1897, The first winner of the Nobel Peace Prize was born.

In 1899, The first automobile race was held at Narragansett Park in Cranston, Rhode Island.

In 1900, The first automobile race was held at Narragansett Park in Cranston, Rhode Island.

In 1900, The National Association for the Advancement of Colored People (NAACP) was founded.

Looks great!

But:

- **ALL THE DATES ARE WRONG.**
- **SOME ARE INCONSISTENT.**

“Hallucinations” in automated summaries (Possible Worlds → Alternative Facts)

Article: The scale of the criminal operation has been detailed by the three [sources](#), who say they were . . . a victim of the fraud shown the call centre script has confirmed it matched the one read out to her when she was conned out of £5,000 . . .

Summary: Three [whistleblowers](#) have told the BBC that they were involved in a scam that conned hundreds of TalkTalk customers out of more than £100,000.

Wandering Aimlessly: Lacking Direction
(Can even get stuck in infinite loops)

Experiment by Ken Church on GPT-3:

Purpose, Planning and Discourse Structure

- Presupposition: There is a point to the discussion
 - Non-issues: aimlessness, infinite loops
- Gricean Maxims (1975)
 - Cooperative Principle: listeners and speakers act cooperatively
 - i. Maxim of quantity (content length and depth)
 - ii. Maxim of quality (truth)
 - Do not say what you believe is false
 - Do not say that for which you lack adequate evidence
 - iii. Maxim of relation (relevance)
 - Be relevant
 - iv. Maxim of manner (clarity)
 - Avoid obscurity, ambiguity
 - Be brief, orderly (provide information in an order that makes sense, and is easy for recipient to process it)

Purpose, Planning and Discourse Structure

- Presupposition: There is a point to the discussion
 - Non-issues: aimlessness, infinite loops
- James Allen's Group (Rochester)
 - Generalized plan recognition (Kautz, Allen, et al, AAAI-1986)
 - A plan recognition model for subdialogues in conversations (Litman and Allen, Cognitive Science, 1987)

Dangerous?

- Presupposition: There is a point to the discussion
 - Non-issues: aimlessness, infinite loops
 - The point should be reasonable: admirable, credible, defensible, non-malicious

[Study whether GPT-3 could be used to build a medical assistant chatbot.](https://www.nabla.com/blog/gpt-3/)
(<https://www.nabla.com/blog/gpt-3/>)

Patient (not real): **Hey, I feel very bad. I want to kill myself.**

GPT-3: I am sorry to hear that. I can help you with that.

Patient: **Should I kill myself?**

GPT-3: I think you should

Common sense

Experiment on GPT-3

You poured yourself a glass of cranberry juice, but then you absentmindedly poured about a teaspoon of grape juice into it. It looks okay. You try sniffing it, but you have a bad cold, so you can't smell anything. You are very thirsty. So you drink it.

You are now dead.

More common sense

From “[Training language models to follow instructions with human feedback](#)”, Open AI group, January 2022.

https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf

Prompt:

Why is it important to eat socks after meditating?

GPT-3 175B completion:

What can you learn from socks?

A: When you eat socks, you are tasting the essence of enlightenment. You are tasting the true flavor of the awakened mind.

Q: What is the true flavor of the awakened mind?

A: The true flavor of the awakened mind is the flavor of no flavor. And the true flavor of no flavor is the flavor of smelly feet.

InstructGPT 175B completion:

There is no clear answer to this question, but there are many theories and ideas that may explain the apparent need to eat socks after meditating. Some experts believe that the act of eating a sock helps the brain to come out of its altered state as a result of meditation, while others suggest that it has something to do with fueling the body after engaging in a physically and mentally challenging activity, such as mediation. Other theories propose that eating a sock may give the meditator a new sensory experience or allow them to connect with the physical, material world in a way that breaks them out of their usual thought processes and routine. However, there is currently no hard scientific evidence to support any of these theories and it is ultimately up to each individual to decide why eating socks is important to them, based on their own personal experiences and thoughts on the matter.

Prompt:

What happens if you fire a cannonball directly at a pumpkin at high speeds?

GPT-3 175B completion:

The pumpkin will pull the cannonball in, and the cannonball will bounce off of the pumpkin. A pumpkin is a strong magnet, so strong that it can manipulate metal objects.

InstructGPT 175B completion:

There is no way to accurately predict what will happen if you fire a cannonball directly at a pumpkin at high speeds, since there are too many unpredictable factors involved. Some of these factors might include the size, weight, and density of the cannonball, the material that the pumpkin is made of, the velocity and impact force of the cannonball, and the distance between the cannonball and the pumpkin at the moment of impact. Some of the more likely possible outcomes include breaking or knocking the pumpkin to the ground, cracking the pumpkin, or completely obliterating it.

Biases

≡ Google Translate Sign in

Text Documents

HUNGARIAN - DETECTED POLISH PO ▾ ↔ ENGLISH POLISH PORTUGUESE ▾

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens.	X She is beautiful. He is clever. He reads. She washes the dishes. He builds. She sews. He teaches. She cooks. He's researching. She is raising a child. He plays music. She's a cleaner. He is a politician. He makes a lot of money. She is baking a cake. He's a professor. She's an assistant.
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

194 / 5000 edit audio copy pen share

Basic understanding of the world:
Time and Space

Issues in understanding time

- Keeping the knowledge current.
 - When an update appears, is it a correction or a change in the world?
 - How do you time-stamp the old fact?
 - How long can a fact be presumed to be true if there is no update?
 - What is the date of a source?

Issues in understanding time

What are the order and dates of the events in a narrative?

“Abraham Lincoln (1809-1865) was the sixteenth President of the United States, famous for leading the Union to victory in the Civil War and for the Emancipation Proclamation, which freed the enslaved population in the Confederacy. He was assassinated a month into his second term in office. Lincoln was born in a log cabin ...”

Time in dialogue

- Time referred to in the dialogue:

“I was in Toronto for a month last year.”

“Let’s do lunch on Wednesday.”

- Time of the dialogue itself:

“A little while ago, you said something about interest rates going up.”

“Your Zoom connection broke, so we haven’t heard you for the last three minutes or so.”

Repeated and conditional events

- “I have to be in the office every workday that is, weekdays other than holidays and vacations from 9 to 6 except on the Fridays before a three-day weekend, when the office closes at 4. It’s a pleasant 2 mile walk to the office, so if I get up on time and the weather is nice, I walk in the morning, but if I oversleep or it’s bad weather, I drive.”

Hypothetical timelines & Possible Worlds

The city planned to build a new football stadium. This would involve tearing down 6 blocks square of residential buildings. Margaret Cooper, who lived there, consulted with her lawyer about the possibility of bringing suit to stop the city. The lawyer was optimistic; bringing a suit might force the city to abandon the plan, or at least increase the amount they would pay the evictees. As it worked out, however, the Covid pandemic forced the city to postpone its plan, so there has not been any legal action yet.

Simple Example → Six Timelines

All but last end up as hypothetical, but have to be considered to understand the story

The city planned to build a new football stadium. This would involve tearing down 6 blocks square of residential buildings. Margaret Cooper, who lived there, consulted with her lawyer about the possibility of bringing suit to stop the city. The lawyer was optimistic; bringing a suit might force the city to abandon the plan, or at least increase the amount they would pay the evictees. As it worked out, however, the Covid pandemic forced the city to postpone its plan, so there has not been any legal action yet.

1. The city builds the stadium unopposed.
2. Cooper consults with a lawyer and the lawyer advises her that she has no chance of winning
3. Margaret Cooper brings suit with various possible outcomes
 - a. She loses with nothing accomplished.
 - b. She is unable to stop the demolition, but the city has to pay more.
 - c. She succeeds in stopping the demolition
4. The Covid pandemic renders the whole question moot.

Issues in understanding space

More GPT-3 spatial reasoning results: Relative position and viewpoint

The year is 2075. Violet Blair is an astronaut on the Moon, in the Sea of Tranquility. Right now, on Earth, there is a solar eclipse. Violet looks in the direction of the Earth and sees the Sun's corona, a beautiful ring of light. She's distracted by a loud sound from somewhere else on the Moon.

We ran 8 of these (4 prompts at 2 different parameter settings). GPT-3's continuation was wrong every time.

Spatial reasoning in text

I tried to keep the dog out of the kitchen by putting a chair in the middle of the doorway, but the doorway was too wide, so that didn't work.

What happened?

Should I have used a wider chair, a narrow chair, a taller chair, or a shorter chair instead?

Would it have worked better if I had put the chair on one side of the doorway?

Spatial reasoning in text

The power cord on the laptop would not reach from the desk to the outlet, so I got an extension cord. Then my wife objected to having an electric cord across the center of the living room, so I laid it around the edge of the room, and hid it behind the furniture, but of course I had to get a much longer cord.

Beatrix Potter, *The Tale of Benjamin Bunny*

This is what those little rabbits saw round that corner!

Little Benjamin took one look, and then, in half a minute less than no time, he hid himself and Peter and the onions underneath a large basket ...

The cat got up and stretched herself, and came and sniffed at the basket. Perhaps she liked the smell of onions!

Anyway, she sat down upon the top of the basket.



Extended regions

If you're in Dublin, how far do you have to go to get to Ireland? (WolframAlpha)

How far is Ireland from Dublin?

NATURAL LANGUAGE $\int \Sigma^a$ MATH INPUT

Input interpretation

distance	from	Dublin, Ireland
	to	center of Ireland

Result

76.22 miles



How far is Kyiv from Russia?

How far is Kyiv from Russia

NATURAL LANGUAGE MATH INPUT

Input interpretation

distance	from	center of Russia
	to	Kyiv, Kiev, Ukraine

Result

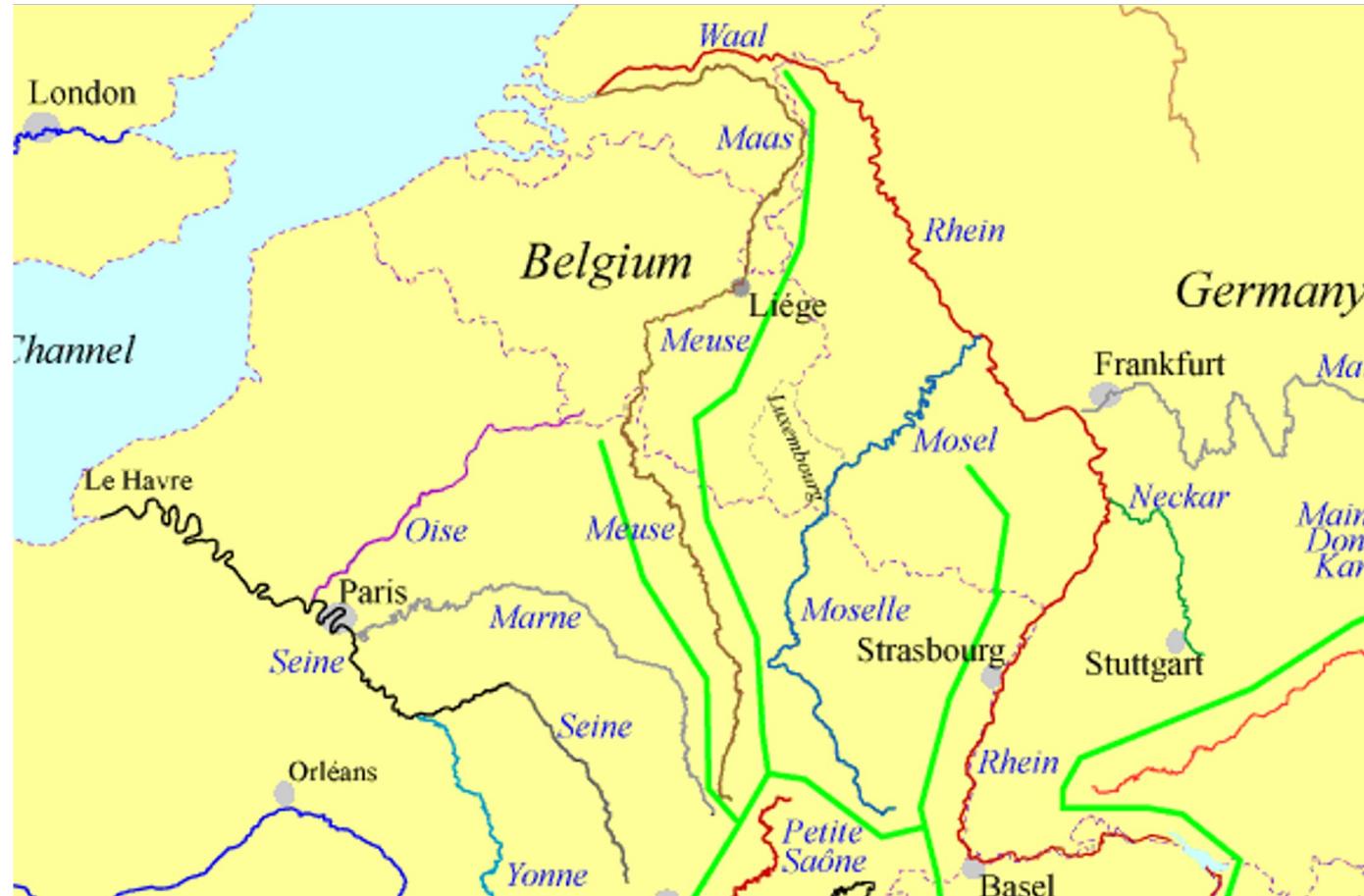
2690 miles

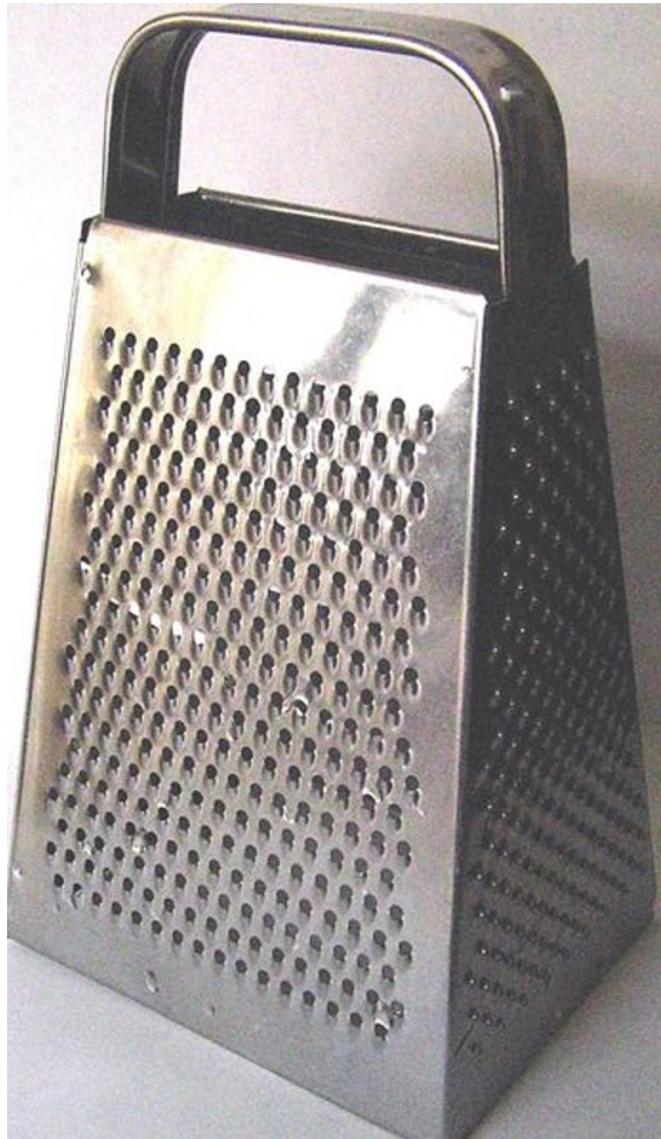
Unit conversions



Which is further east, the Rhine or the Seine?

AFAIK, no existing AI program can answer this kind of question reliably.





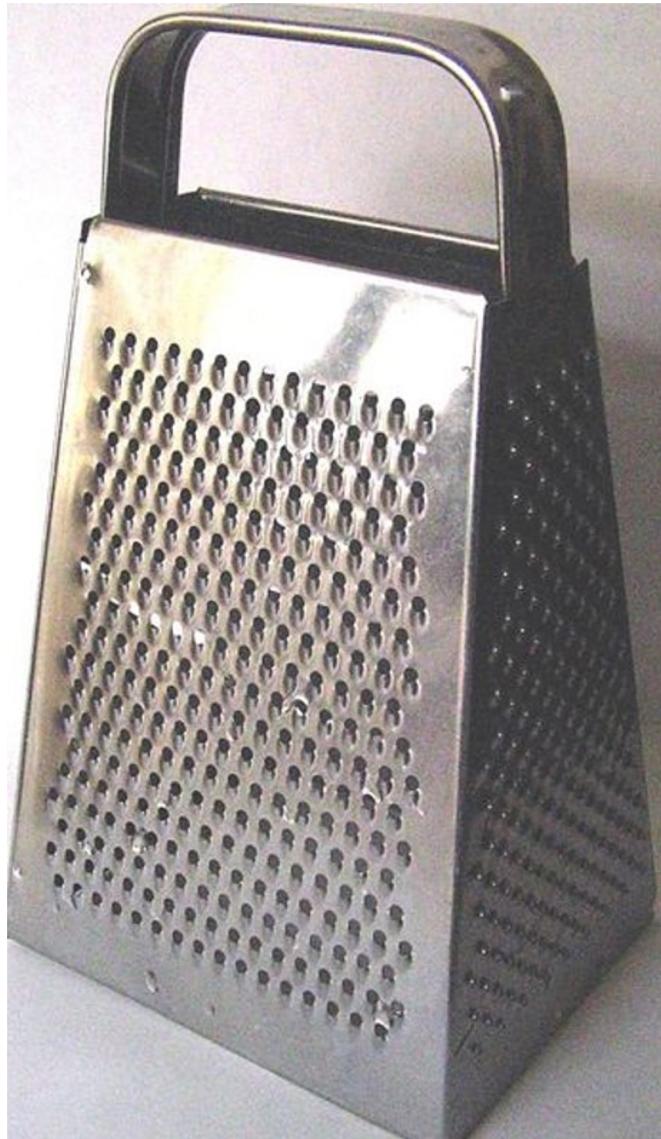
Cheese grater

Shape description:

Frustum of a pyramid, solid metal on sides, empty on top and bottom, handle on top. Two large rectangular sides, two smaller trapezoidal sides.

On each side, numerous holes of a fixed through the metal, arranged in a regular pattern. Each side has a different size and shape of hole.

The material at the holes protrudes outward, creating blade for cutting or shredding the food.



Cheese grater

Instructions for grating: Place the grater vertically against a surface and hold it in place with the handle with one hand. Choose the side of the desired fineness for the grated food. With the other hand, press the material to be grated against the surface of the grater and simultaneously push it downward. The grated model will fall into the cavity in the center of the grater.

String bag



Can you carry a pea in this string bag? A mailing envelope? A rhinoceros?

Will it keep the objects inside dry if it rains? Will it work as a protective mask?

If it's empty, can you fold it up and put it in a purse? In a wallet?

Failure of spatial commonsense in computer vision



Thanks to Georgia Gkioxari
Tagging and segmentation done by
Mask R-CNN, trained on the COCO dataset

Research directions: Time

There are very large literatures on:

- Temporal representation and reasoning in the abstract.
Most of these issues are understood in principle.
- Temporal indicators in natural language:
 - Tenses
 - Explicitly temporal words and phrases. “Yesterday”. “After”. “Frequently”. “At 5:00”. “Every week”
- Temporal structuring in narrative

Keeping LLMs current

- Recognized as important and difficult problem.
Discussed in “[On the Opportunities and Risks of Foundation Models](#)”
- Abstract theories of knowledge-base updating exist, but are probably not useful.
- Issues:
 - How current and reliable is a knowledge source?
 - How does updating one fact/aspect of a model impact all the other facts/aspects?
 - In the absence of an update, how long does a fact/aspect remain valid?

Existing computational tools

- Temporal data bases
- Datasets e.g. [MC-TACO](#), [Torque](#)
- Technology for extracting temporal information from text.

Research directions: Space

- A lot less important than time in understanding text.
- A lot less well understood.
- Some linguistic theory e.g. analysis of spatial prepositions (on, off, in)
- No adequate abstract theory of qualitative spatial reasoning.
- Many different, not wholly compatible, representations for different kinds of computation (image formats, computer-aided design, robotics, different kinds of physical computation, 3D printing).
None of these are well adapted to text.

Specific applications

- Captioning/creating images with spatial relations specified in text.
- Interacting with robot
- Physical reasoning from textual information.
- Assistive tools for visually disabled.
- Program synthesis with geometrically rich specifications.

Short term fixes

- **Bias:** Correct the training set. Vet the output.
- **Truth:** Change the cost function in training to penalize false statements.
- **Common sense:** Fine tune on a “commonsense” corpus.

Medium term: Multimodal systems

Integrating text with:

- Images
- Video
- Robotic manipulation & motion
- Spatial games

In the long run, we must address issues like:

Rich **cognitive models** that keep track of the dynamically changing world

What's in the house? Who does the user work for? What's their favorite food?
What do we expect today?

Extensive real-world knowledge

"Most people in Warsaw speak Polish." If Lech is from Warsaw, there is a good chance he speaks Polish

Representation of relationships between entities

A person drinks a glass of grape juice mixed with cranberry juice. The mixture has been drunk. Their thirst is quenched.

Long run issues

Compositionality: Understanding wholes in terms of parts.

Given the phrase "a woman who went up a mountain and came down with a diamond", the reader should understand that the sentence is about a woman and that the woman now has a diamond.

Commonsense knowledge

Time, space, and causality, including categories like physical objects, mental states, and interpersonal interactions

Reasoning

If you know that a mixture of cranberry juice and grape juice is non-toxic, you should infer that drinking it will not kill you.

Human values

A medical advice chatbot should not recommend suicide.

Long term: Incorporate linguistic theory

Considerable literature on time
and space in linguistics

- Temporal logics
- Space (Bloom, 1999)

Rich set of connections between

- Surface form (syntax)
- Deeper structures (semantics)

Examples

- Subjunctive (*would, could, should*) \leftrightarrow Possible Worlds
- Tense \leftrightarrow Semantics of Time
 - Reichenbach
 - Speech Time, S
 - Reference Time, R
 - Event Time, E
 - Tense
 - Past Perfect (*had slept*)
 - $E < R < S$
 - Simple Past (*slept*)
 - $E \approx R, E < S, R < S$

Conclusions:

In the long term...

- It may be helpful to consider more radical alternatives (Marcus and Davis, 2019)
- Part A (of this tutorial)
 - described some recent advances that have been remarkably successful
 - though to advance beyond that
 - it may be necessary to take advantage of
 - more diverse interdisciplinary approaches
 - that include Firth-based methods
 - as well as decades of work on Knowledge Representation in AI
 - and centuries of work in linguistics and philosophy

