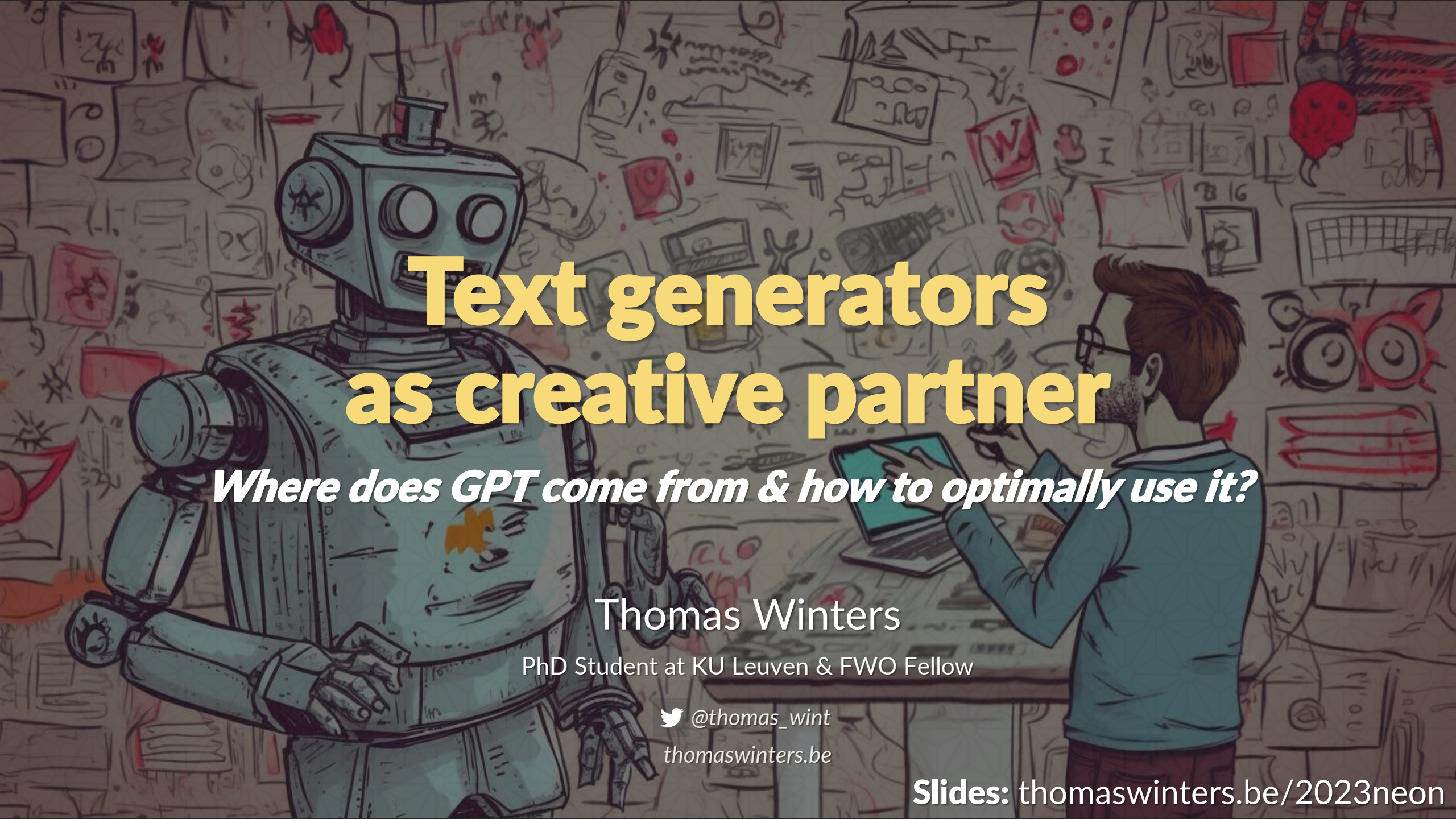


Text generators as creative partner

Where does GPT come from & how to optimally use it?

Thomas Winters

PhD Student at KU Leuven & FWO Fellow

@thomas_wint

thomaswinters.be

Slides: thomaswinters.be/2023neon

Who has used a GPT system before?



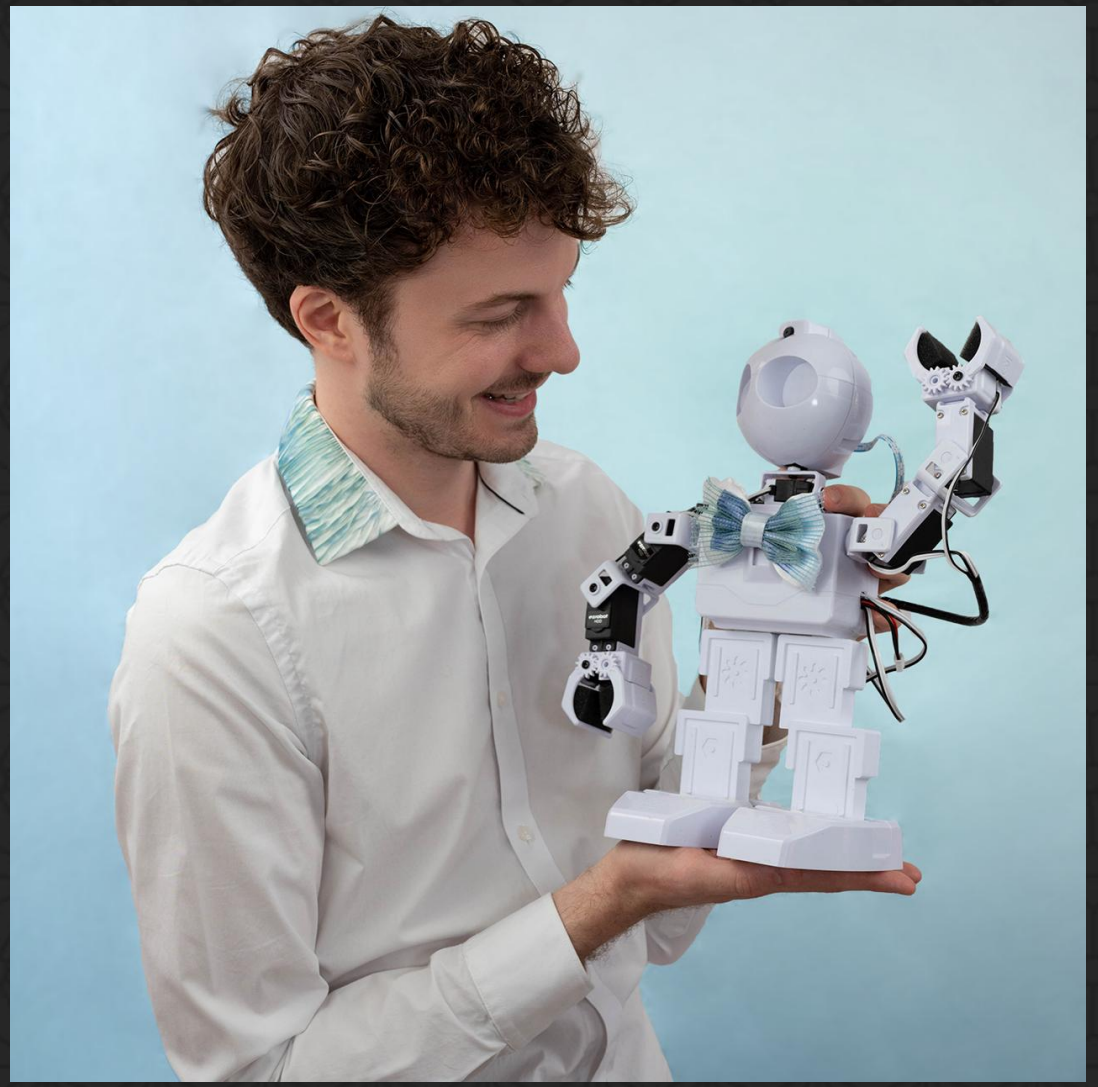
Who has used GPT to solve an actual need (work, study...)?



Thomas Winters

- PhD student in creative AI & language models since 2018 at DTAI
- Co-authored RobBERT model
  (state-of-the-art Dutch BERT language model)
- Since GPT-2 (2019) experience with met GPT-modellen *(humor, Improbotics...)*

KU LEUVEN LEUVEN.AI DTAI



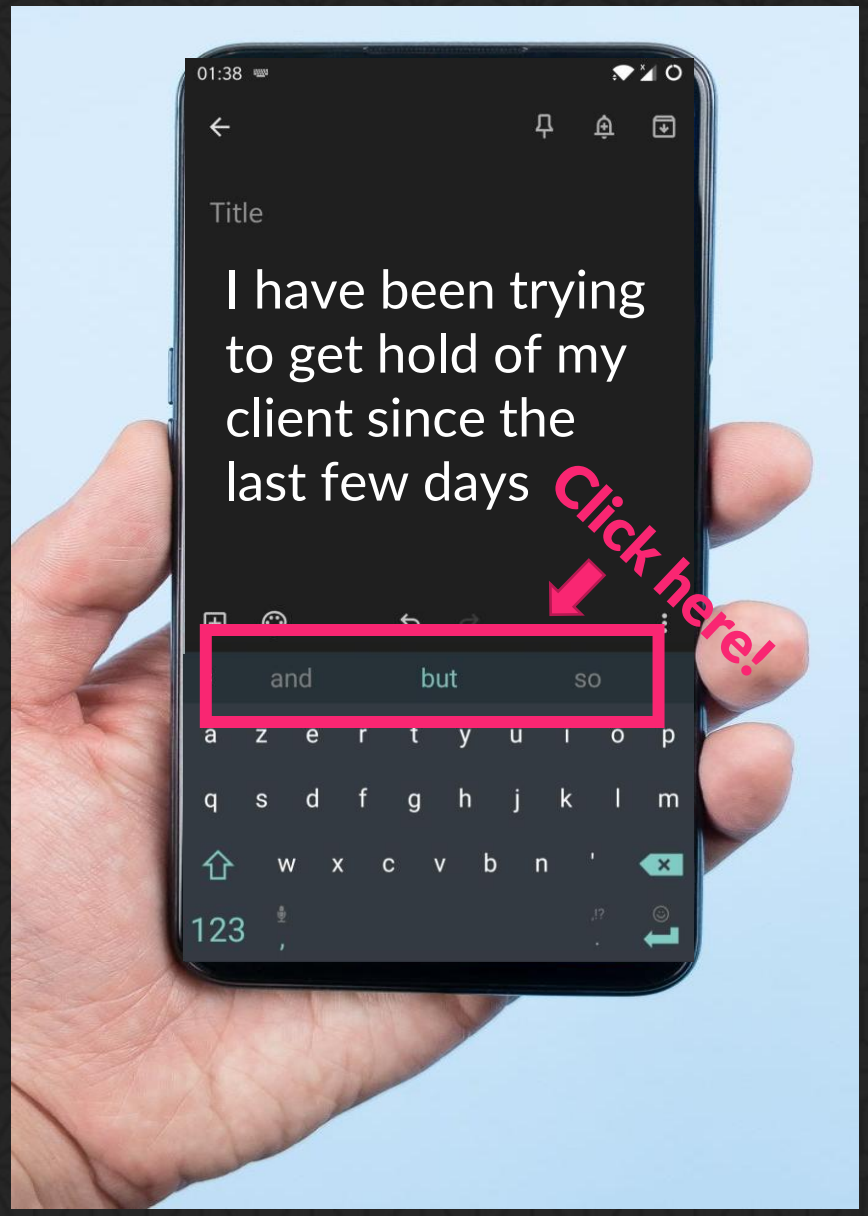


Textgeneration History

Overview on the ~4 GPT waves

Part 1

How do autoregressive language models work?



Demo

1. Open your smartphone keyboard
2. Press on any autocomplete word
3. Press 10-20 times
4. You've just generated a sentence using an AI trained to sound similar to you!

Autocomplete counted how often you used certain words after other words in previously typed texts

Example: TorfsBot

1. Counted in all tweets & columns of Rik Torfs how often one word follows others



2. Then starts with two random words and randomly select words Rik used after the last couple of words





Autoregressive models like GPT



≈ autocomplete on steroids

The next word for this charade is

sentence = 10.48%
list = 6.61%
phrase = 6.43%
definition = 6.12%
sequence = 4.07%
char = 0.15%

Total: -6.49 logprob on 1 tokens
(33.86% probability covered in top 6 logits)

Probability of next token given *previous tokens*

GPT training



Whole textual internet

Training process of several million $

GPT-4

Best public language model

Comment: Toledo is so slow and unstructured, I can't find anything!
Sentiment: Negative

Comment: The new Ultra seems nice!
Sentiment: Positive

Comment: I keep getting lost trying to find Bethlehem aula!
Sentiment: Negative

Comment: I love the winter break!
Sentiment: Positive

Comment: i can't find the athletic gyms during det period!
Sentiment: Negative

Comment: I gotta take to long detour to the bathroom!
Sentiment: Negative

Comment: Parking is a bit tricky to on a weekday!
Sentiment: Negative

First GPT-3 (2020)

Can do task given few examples
= **Pretrain & Few-shot Prompt**

Keeps babbling after answer

e.g., "hallucinating" non-existing comments

4000 tokens input-output limit

Write a list of reason's why it's great to work at KU Leuven.

1. Highly competitive salaries and benefits packages.
2. Specialized research opportunities.
3. Advanced and cutting-edge research facilities.
4. A diverse and welcoming atmosphere.
5. Access to global contacts and networking.
6. High career mobility.
7. Prestigious international reputation.
8. Abundant network of experts in multiple fields.
9. Global recognition of academic excellence.
10. Encouragement of entrepreneurship and innovative initiatives.

Instruct GPT-3 (2022)

Follows instructions
= **Pretrain & Zero-shot Prompt**

Shuts up when completed

Less toxicity & hallucinations

RLHF *(Reinforcement Learning from Human Feedback)*

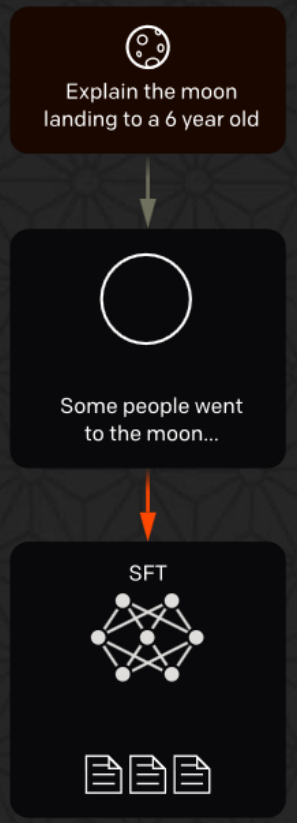
Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3 with supervised learning.



Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



ChatGPT

TH

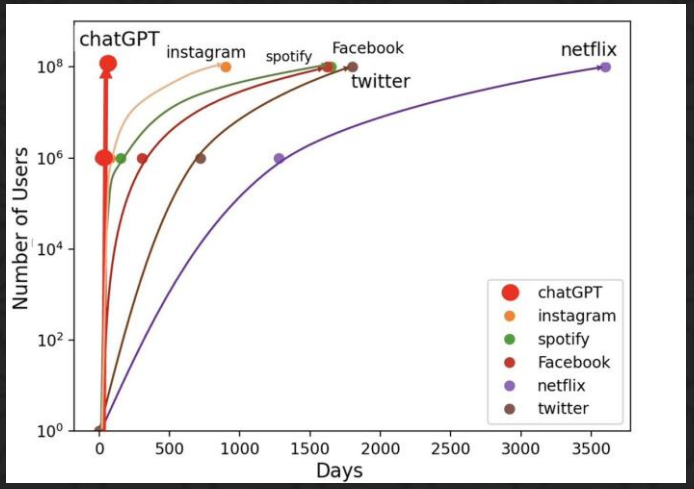
Write a list of titles for a talk about the history of large language models, chain-of-thought prompting, let's think step by step and continuous prompting

1. "Tracing the Evolution of Large Language Models"
2. "Chain-of-Thought Prompting: Understanding the History and Development"
3. "Step by Step: Examining the Advancements in Language Modeling"
4. "Continuous Prompting: A Journey Through the History of Language Processing"
5. "From ELMO to GPT-3: The Rise of Large Language Models"

Uses RLHF-based GPT-3

Finetuned on chats

Fastest growing platform ever



Tokens	Characters
119	414

GPT-3 tokenizers words to be able to input them as integers in the model.

The tokens are chosen in such a way that common substrings have their own token. It is purely frequency-based, and do not need to make morphological sense: e.g. gothic

In English, common words are usually a single token.
In het Nederlands is dat vaak andere koek.
Vandaar dat het moeilijker lettergrepen en rijmen snapt in het Nederlands.

https://platform.openai.com/tokenizer

Main strengths of GPT

USER

Paraphrase the mail below in extremely formal style

Hi Veerle! Yes, I'll be able to give a talk about the inner workings of AI text generators & prompt engineering for your science communication training day! -Thomas

ASSISTANT

Dear Ms. Veerle,

I am writing to inform you that I shall indeed be available to orate on the subject of the intricate operations intrinsic to AI text generators and the process of prompt engineering. I would be delighted to share my insights during your forthcoming science communication training day.

Yours faithfully,

Mr. Thomas

USER

Rewrite it in extremely simple language so a 5-year-old child understands it

ASSISTANT

Hey Veerle! Yup, I can come and chat about how cool robots make up words for your special school day!
From,
Thomas

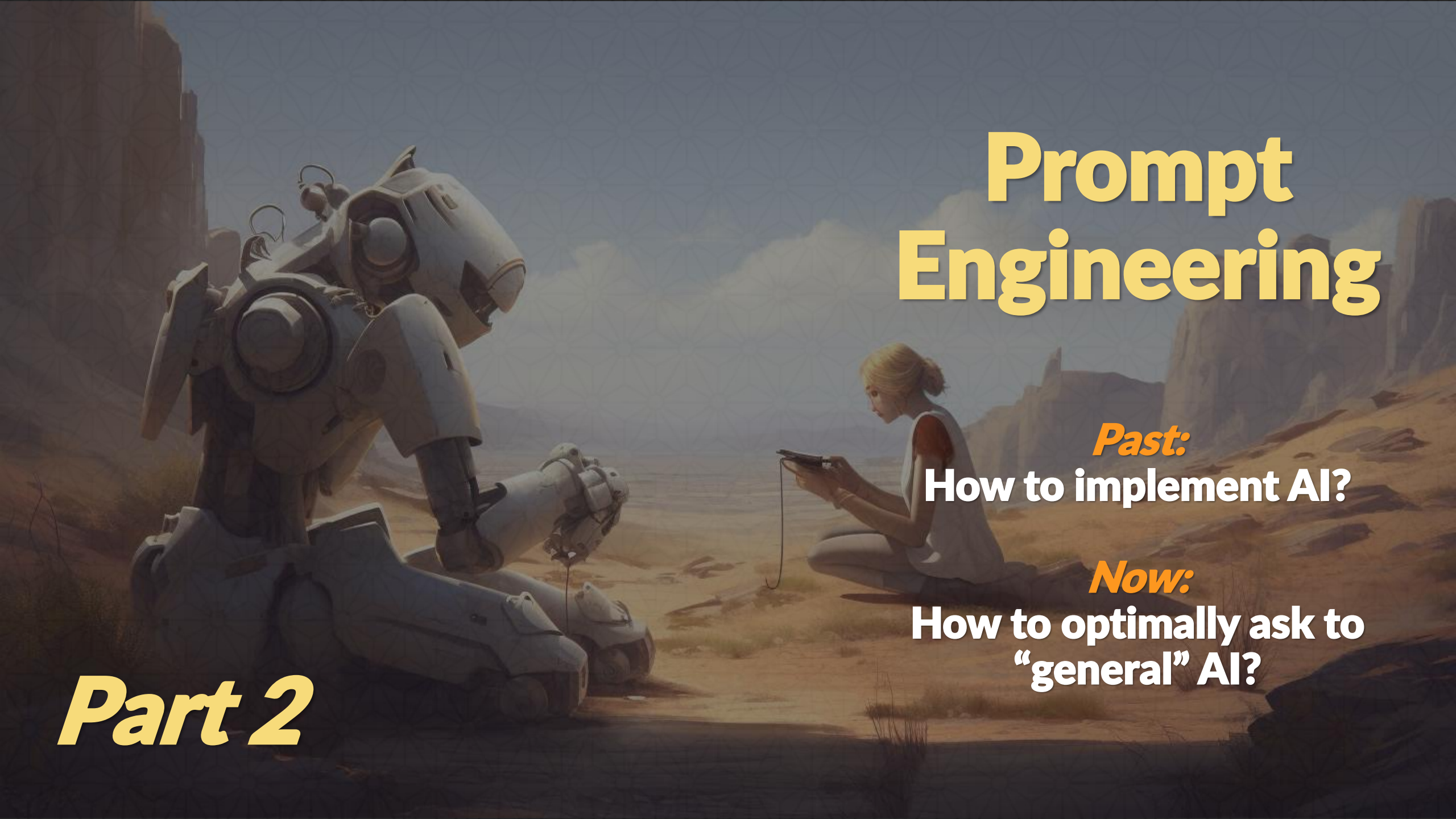
Copy writing → Copy editing

Amazing at paraphrasing to any style & format

Even to extremer styles

USER	Rewrite the mail to sound like a pirate wrote it, full of references to pirate life and the sea
ASSISTANT	Ahoy there, Veerle! Aye, I be ready to spin a yarn 'bout the mystery heart o' them word-makin' sea monsters called AI, and how we be steerin' their prompts, fer your gatherin' of learned sea dogs! Cheers, Cap'n Thomas

USER	Rewrite it as a thrilling Disney story plot
ASSISTANT	Once upon a time, there was our adventurous hero, Thomas. On one bright day, he received a magical message from the wise princess Veerle. She was assembling a grand celebration, a day of knowledge sharing, in her magnificent kingdom of science communication. Full of courage, Thomas pledged, "Indeed, Princess Veerle, I am prepared to embark on a remarkable journey. I will dive into the mysterious world of the magical creatures known as AI text generators, unravel their intricate inner workings, and reveal the delicate art of guiding their prompts. On your grand day of scientific enlightenment, I shall narrate this wondrous tale." And so, the stage is set for an epic tale of technology and wizardry in the enchanting kingdom of science communication. Don't miss Thomas's thrilling saga on the much-anticipated day!



Prompt Engineering

Past:

How to implement AI?

Now:

How to optimally ask to "general" AI?

Part 2

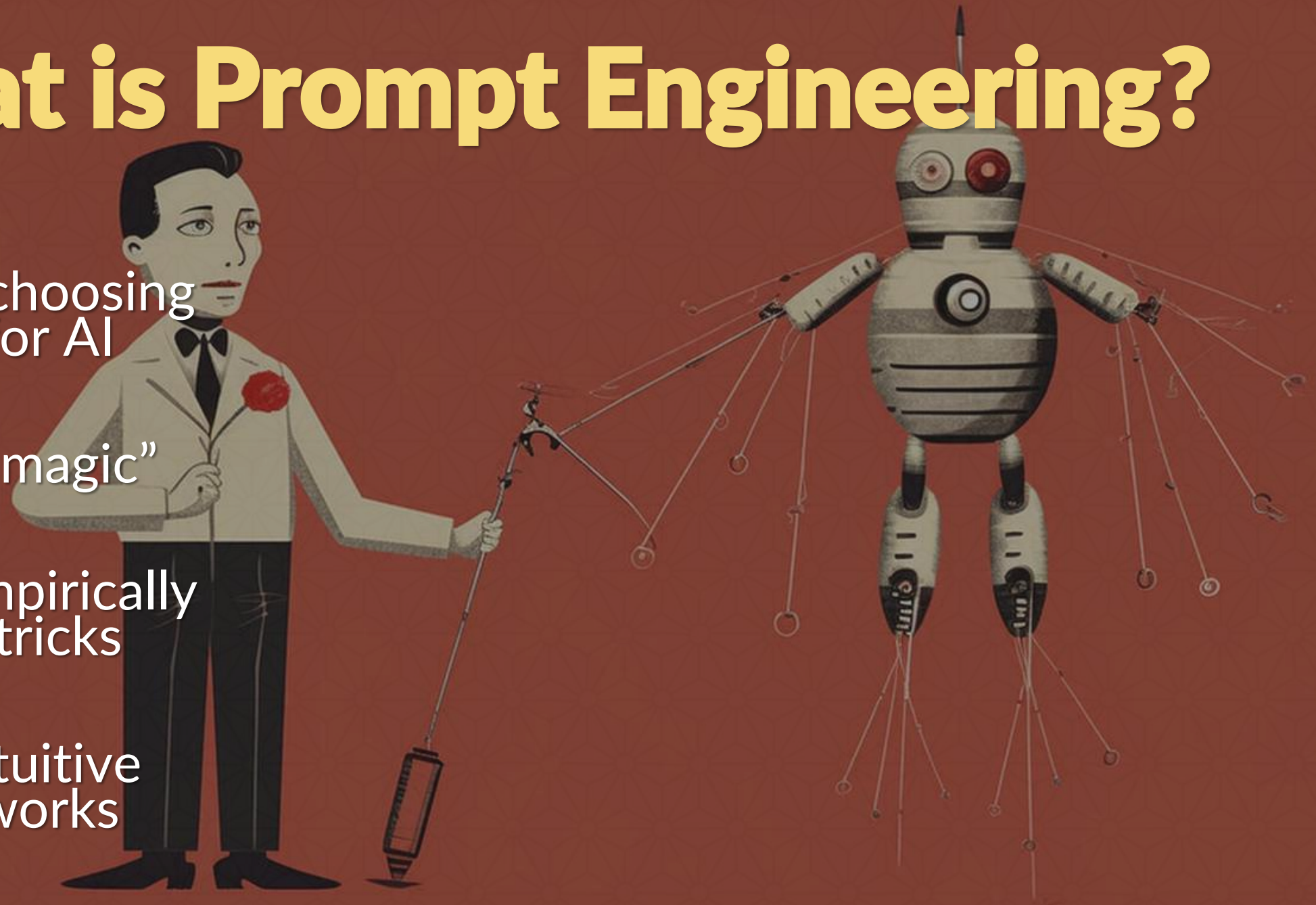
What is Prompt Engineering?

Optimally choosing words for AI

"Voodoo magic"

But also empirically proven tricks

Often intuitive why it works




6 tricks for enhancing
any creative task prompt

Base: simple write instruction prompt

USER	Write a joke for literature students
ASSISTANT	Why did Shakespeare make terrible coffee? Because he always put too much Hamlet and not enough Macbrew!

Trick 1: Give handful of examples *("Few-shot")*

USER

Write list of jokes for literature students
- Why did Shakespeare write in pen? Because pencils confused him: 2B or not 2B?
- I've taken up speed reading. Last night, I read Harry Potter in 20 seconds. I know it's only 2 words, but it's a start.
- Studying about fairy tales can be a Grimm business!

ASSISTANT

- Why did the grammarian go to jail? Because he committed a sentence.
- Why don't scientists trust atoms? Because they make up everything.
- What do you call a group of cows reading a book together? A herd of intellectuals.
- Why did the protagonist break up with the author? Because he was always putting words in her mouth.
- Why did the tomato turn red? Because it saw the salad dressing.
- Why did the poet break up with his girlfriend? Because she was always a syllable short of a haiku.
- What's the difference between a poorly dressed man on a trampoline and a well-dressed man on a trampoline? Attire.
- What do you call a book club that's been stuck on the same novel for years? Groundhog Day.
- What did Jane Austen say when she was offered a second cup of tea? "I must decline, for I am already Austen-tea-fied."

Trick 2: Explain role *("role-prompting")*

USER

You are a world-renowned expert in comedy with a proven track-record in writing jokes and stand-up comedy for audiences of students all around the world. I am looking for jokes that would appeal literature students. Your task is to write a list of literature-related jokes that relate to student life.

ASSISTANT

1. Why did the literature student always bring a pillow to class? Because they were studying the "dream" sequence.

2. How do literature students stay cool in summer? They sit under their Poe-tree.

3. Why did the literature student refuse to put their initials on their paper? They didn't want to be accused of self-Plath.

4. What's the difference between an English major and a pizza? At least a pizza can feed a family of four!

Tell ChatGPT to pretend to be an expert.

For example:
You are an expert X with proven track record on X'. I am looking to Y. Your task is Z. My first command is …

"localizes" task in model

More inspirations
https://github.com/f/awesome-chatgpt-prompts

GPT is bad at math?

Thomas takes a train from Leuven to Gent. Leuven Station and Gent have a distance of 100km. The train travels at 50km/h. The train also stops three times for 10 minutes each. How long is the train ride?

2 hours

Trick 3: Ask for reasoning steps *("Zero-shot chain-of-thought")*

Thomas takes a train from Leuven to Gent. Leuven Station and Gent have a distance of 100km. The train travels at 50km/h. The train also stops three times for 10 minutes each. How long is the train ride? Let's think step by step.

The train travels at 50km/h, so it will take 2 hours to travel 100km. However, the train stops three times for 10 minutes each. This means that the train will spend a total of 30 minutes stopped. Therefore, the total train ride will be 2 hours and 30 minutes.

USER

You are an expert comedy-writer with a proven track-record for writing jokes for students.
Every time you write a joke, you draft the joke by going through several distinct steps.

Q: Your task is write jokes that appeal to literature students

A: Let's think step by step.

Works for any task!

ASSISTANT

Step 1: Identify the target audience and their interests
- Target Audience: Literature students
- Interests: Classic literature, famous authors, literary terms and techniques, wordplay, character analysis, writing process

Step 2: Brainstorm relevant topics and concepts
- Shakespeare
- Jane Austen
- Edgar Allan Poe
- Charles Dickens
- Plot twists
- Unreliable narrators
- Poetry vs. prose
- Metaphors and similes
- Writer's block
- Allusions and references

Step 3: Develop joke ideas based on the brainstormed topics
- Shakespeare's love for puns
- Jane Austen's witty social commentary
- Edgar Allan Poe's dark humor
- Charles Dickens' character names
- Unexpected plot twists in literature
- Unreliable narrators causing confusion
- Comparing poetry to prose

Best in format "Q: ....
A: Let's think step by step."

Step 3: Develop joke ideas based on the brainstormed topics
- Shakespeare's love for puns
- Jane Austen's witty social commentary
- Edgar Allan Poe's dark humor
- Charles Dickens' character names
- Unexpected plot twists in literature
- Unreliable narrators causing confusion
- Comparing poetry to prose
- Misinterpreting metaphors and similes
- Writer's block struggles
- Allusions to other literature in everyday conversations

Step 4: Write jokes based on the developed ideas

1. Why did Shakespeare make terrible friends? Because he always had to have the last (word)play.

2. Jane Austen walks into a bar and orders a drink. The bartender says, "Austentatious choice, my lady."

3. How does Edgar Allan Poe like his coffee? Dark, just like his sense of humor.

4. Why was Charles Dickens always invited to parties? He was great at creating characters.

5. Why do literature students love plot twists? They keep us in suspense, just like our grades.

6. What did the unreliable narrator say to their therapist? "I promise, I'm not making this up!"

7. What's the difference between poetry and prose? Poetry is when you spill your feelings on the page, and prose is when you clean it up.

8. Why did the metaphor get in trouble? It was caught simile-ing.

9. What's a writer's favorite type of block? Writer's block, of course – it's the only one they can't avoid.

Step 5: Edit and revise jokes for clarity and punchiness

1. Why did Shakespeare have few friends? He couldn't resist a (word)play on words.

2. Jane Austen walks into a bar and orders a drink. The bartender smirks, "How austentatious!"

3. How does Edgar Allan Poe take his coffee? Dark and brooding, like his stories.

4. Why was Charles Dickens the life of the party? His character creations were a hit!

5. Why do literature students thrive on plot twists? They're used to being kept in suspense.

6. The unreliable narrator in therapy: "Trust me, I'm not making this up!"

7. Poetry vs. prose: Poetry spills feelings on the page; prose wipes it clean.

8. The metaphor got in trouble for simile-ing – it couldn't resist a pun.

9. Writers love one type of block: Writer's block, the one they can't escape

Any “chain of thought”-provoking prompt helps!

USER

You are a world-renowned expert in comedy with a proven track-record in writing jokes and stand-up comedy for audiences of students all around the world. I am looking for jokes that would appeal literature students. Your task is to write a list of literature-related jokes that relate to student life. You write every joke by first thinking step by step.

ASSISTANT

Joke 1:
1. Think about a common theme in literature that students can relate to.
2. Consider the theme of procrastination, which is relatable for students.
3. Relate it to a famous literary figure.

Joke: Why did Shakespeare always seem to wait until the last minute to write his plays? Because he knew that "all's well that ends well."

Trick 4: Make it do multiple explicit steps

USER

To write a satirical joke, you must first complete each of six different steps.

1. topic handles: identify the two main distinct topics that define this article headline.
2. associations: write at least three distinct associations per topic of step 1
3. link: find a funny surprising link between an association from each topic of step 2.
4. punchlines: write three funny punchlines about association link found in step 3
5. best punchline: select the funniest and most original punchline from step 4.
6. full joke: write a joke that first repeats the original news headline from the input, followed by a text linking it to the best punchline from step 5, and then repeats the punchline from step 5 itself.

News headline: Literature student found dead after watching talk on how computers generate humor
1. topic handles:

ASSISTANT Literature student, computers generating humor

2. associations:
Literature student: bookworm, analyzing texts, poetry
Computers generating humor: artificial intelligence, algorithms, failed jokes

3. link: failed jokes causing death

4. punchlines:
a. The computer's humor was so bad, it was literally killer!
b. The student should have stuck to Shakespeare's jokes – at least they're only figuratively dead!
c. Turns out, computers can generate humor so bad, it's to die for.

5. best punchline: Turns out, computers can generate humor so bad, it's to die for.

6. full joke: Did you hear about the literature student who was found dead after watching a talk on how computers generate humor? I guess they experienced firsthand that computers can generate humor so bad, it's to die for.

Trick 5: Reasoning process examples *("Few-shot chain-of-thought")*

Instead of giving examples like

Input: …
Output: …

Add explanation BEFORE output *(so it can be used when calculating the output):*

Input: …
Reasoning: …
Output: …

Trick 5: Reasoning process examples *("Few-shot chain-of-thought")*

Input: vacuum cleaner
Output: I decided to sell my Hoover... Well, it was just collecting dust.

→

Input: vacuum cleaner

Proverb: collecting dust

Reasoning: Let's think step by step. Collecting dust means you don't use something. Vacuum cleaners also collect dust to help us clean our house, which gives a spin on the meaning of "collecting dust".

Punchline: my vacuum cleaner is collecting dust

Joke: I decided to sell my Hoover... Well, it was just collecting dust.

Trick 5: Reasoning process examples *("Few-shot chain-of-thought")*

USER

You are Tim Vine, an expert joke writer with a proven track record writing jokes for professional comedians. I am looking for jokes about given topics. Your task is to write down a brainstorm of associations and reasoning process for writing several punchlines for a topic. Your topics are usually proverbs or words with multiple meanings.

Topic: vacuum cleaner
Proverb: collecting dust
Reasoning: Let's think step by step. Collecting dust means you don't use something. Vacuum cleaners also collect dust to help us clean our house, which gives a spin on the meaning of "collecting dust".
Punchline: my vacuum cleaner is collecting dust
Joke: I decided to sell my Hoover... Well, it was just collecting dust.

Topic: arms
Proverb: serves him right
Reasoning: Let's think step by step. Serving someone right means you got your deserved punishment. Serving could also means performing a duty. So serving someone right means they do something right-handedly.
Punchline: if you miss your left arm, you can only serve right.
Joke: A friend of mine's got a left arm missing. Serves him right.

Topic: literature student

ASSISTANT

Proverb: judge a book by its cover
Reasoning: Let's think step by step. Judging a book by its cover means making superficial judgments. A literature student should know better than to judge books by their covers, as they study the content. We can play on the idea that a literature student might be judged by their appearance, but prove to be more knowledgeable.
Punchline: don't judge a literature student by their cover
Joke: My friend's a literature student, but you wouldn't know it by looking at him. I guess you can't judge a book by its cover – or a literature student!

ASSISTANT

Topic: breaking the ice
Reasoning: Let's think step by step. Breaking the ice means initiating conversation, creating a comfortable atmosphere. Breaking ice can also mean physically breaking ice, like in ice-skating or ice fishing.
Punchline: ice-skaters are great at parties
Joke: Why are ice-skaters always the life of the party? They're great at breaking the ice.

Topic: the early bird catches the worm
Reasoning: Let's think step by step. This proverb means that the person who arrives early or starts early gets the best opportunities. Early birds are morning people, worms are disliked by some people due to their slimy texture.
Punchline: worms must hate mornings
Joke: If the early bird catches the worm, then mornings must be a real nightmare for worms.

Topic: a watched pot never boils
Reasoning: Let's think step by step. This proverb means that time feels slower when you're waiting for something to happen. Watching a pot can imply waiting for water to boil, but it can also mean watching anything that takes time.
Punchline: my kettle hates attention
Joke: My kettle must be really shy because every time I watch it, it refuses to boil.

TH

You are a world-renowned expert in planning scientific & technology-related events, with a proven track record for organising events for the Flemish Department of Economics, Science and Innovation. Your task is to help me come up with ideas for new events, while constantly asking questions to better grasp what I'm looking for. Do you understand what I'm saying?

Yes, I understand your request. As a language model, I can provide you with ideas for scientific and technology-related events and ask questions to better understand your specific needs and requirements. Please feel free to share more details about the type of events you are interested in organizing and any specific goals or objectives you hope to achieve through these events.

TH

I would like to organise an informal and accessible event that demonstrates the opportunities large language models (like GPT-3 and ChatGPT) bring to the Flemish companies

That sounds like an interesting event! Here are some initial ideas and questions to help us better define the scope and format of the event:

1. Would you like to focus on a specific industry or sector in Flemish companies, such as finance, healthcare, or retail, or do you want to showcase the potential benefits of large language models across various industries?
2. How long would you like the event to be, and how many attendees are you expecting?
3. Would you like to invite guest speakers or have panel discussions on the use of large language models in industry?
4. Would you like to showcase live demonstrations of large language models in action and their impact on business outcomes?
5. Would you like to organize any workshops or interactive sessions to help attendees

Trick 6: Force ChatGPT to ask questions

ChatGPT is trained to answer immediately (RLHF)

→ Force it to ask questions!

Best easy prompt is:

You are an expert X with proven track record on X'. I am looking to Y. Your task is Z, while constantly asking questions to better grasp what I'm looking for. Do you understand what I'm saying?

https://www.youtube.com/watch?v=QmA7S2iGBjk



Limitations & Risks

Hallucinations

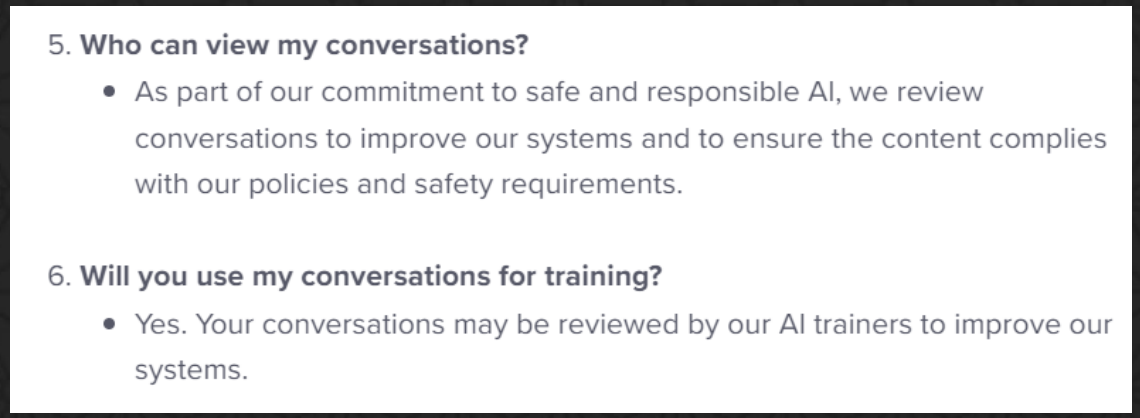
USER	Write a Wikipedia page for TorfsBot
ASSISTANT	TorfsBot is a sophisticated artificial intelligence (AI) bot developed by the OpenAI research group. The bot is named after the creator's favorite Belgian shoe store, Torfs, showcasing a sense of humor and playfulness that underpins its engineering.

GPT will confidently bullshit

Privacy

If online platform: can save your requests!

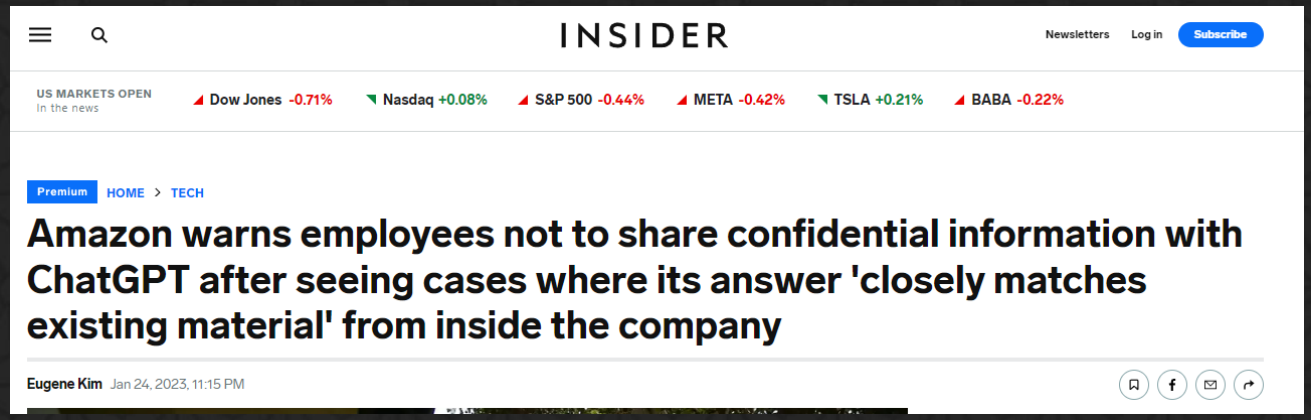
Watch out with sensitive info



5. **Who can view my conversations?**
   - As part of our commitment to safe and responsible AI, we review conversations to improve our systems and to ensure the content complies with our policies and safety requirements.

6. **Will you use my conversations for training?**
   - Yes. Your conversations may be reviewed by our AI trainers to improve our systems.

https://help.openai.com/en/articles/6783457-chatgpt-general-faq



INSIDER

Newsletters Log in Subscribe

US MARKETS OPEN In the news | Dow Jones -0.71% | Nasdaq +0.08% | S&P 500 -0.44% | META -0.42% | TSLA +0.21% | BABA -0.22%

Premium HOME > TECH

Amazon warns employees not to share confidential information with ChatGPT after seeing cases where its answer 'closely matches existing material' from inside the company

Eugene Kim Jan 24, 2023, 11:15 PM

https://businessinsider.com/amazon-chatgpt-openai-warns-employees-not-share-confidential-information-microsoft-2023-1

Limited context window

Can only look back limited number of words

→ **Difficulties with long documents**

GPT-3 / ChatGPT: 4096 tokens (~3K words)

GPT-4: soon 32K tokens (~25K words)

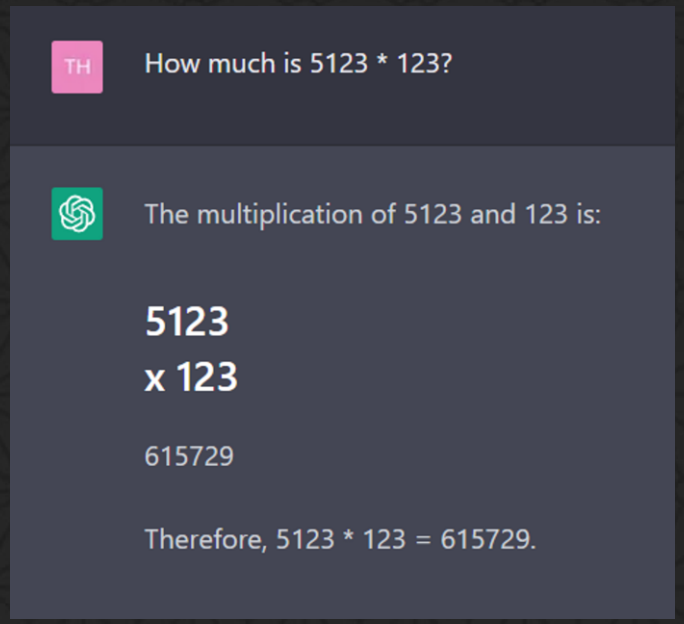
“Solution”: piece-wise document summary

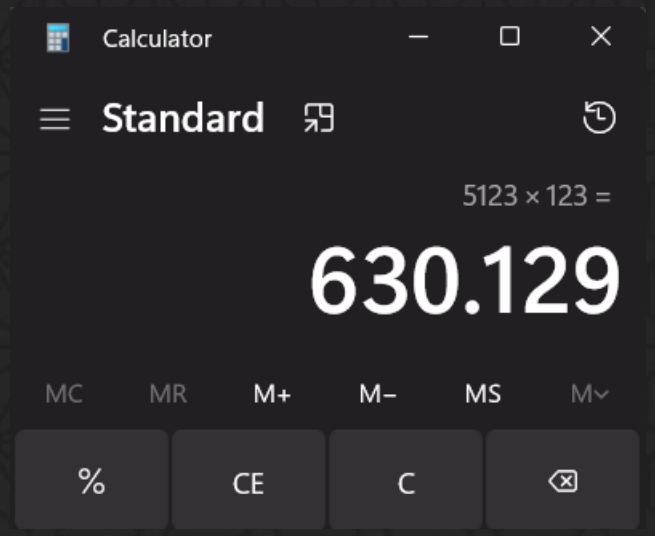
Math

Tokenization problems

The tokens of numbers differ drastically: see 8000, 8001, 8020

& no calculator available!
→ Just next word prediction

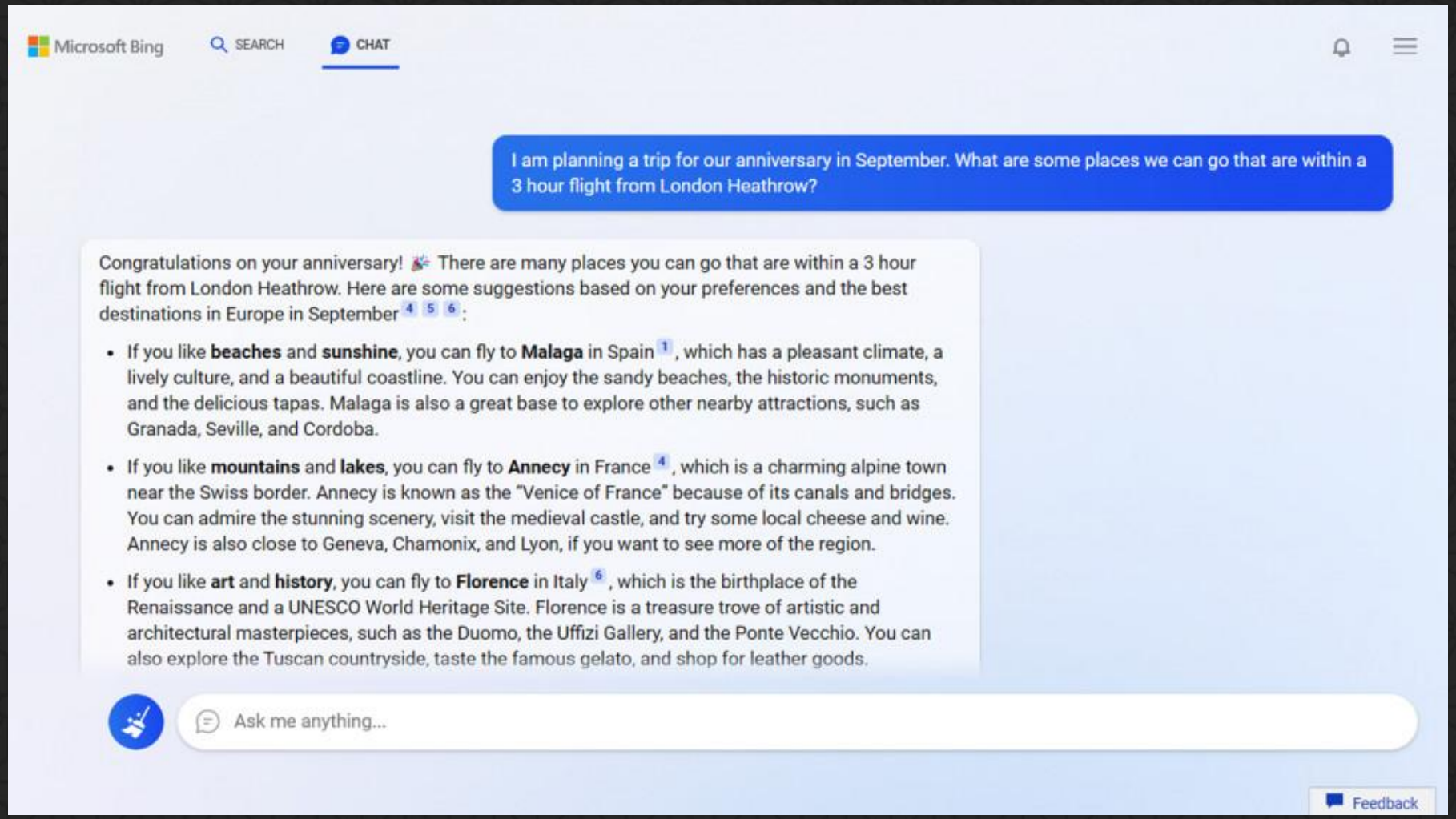






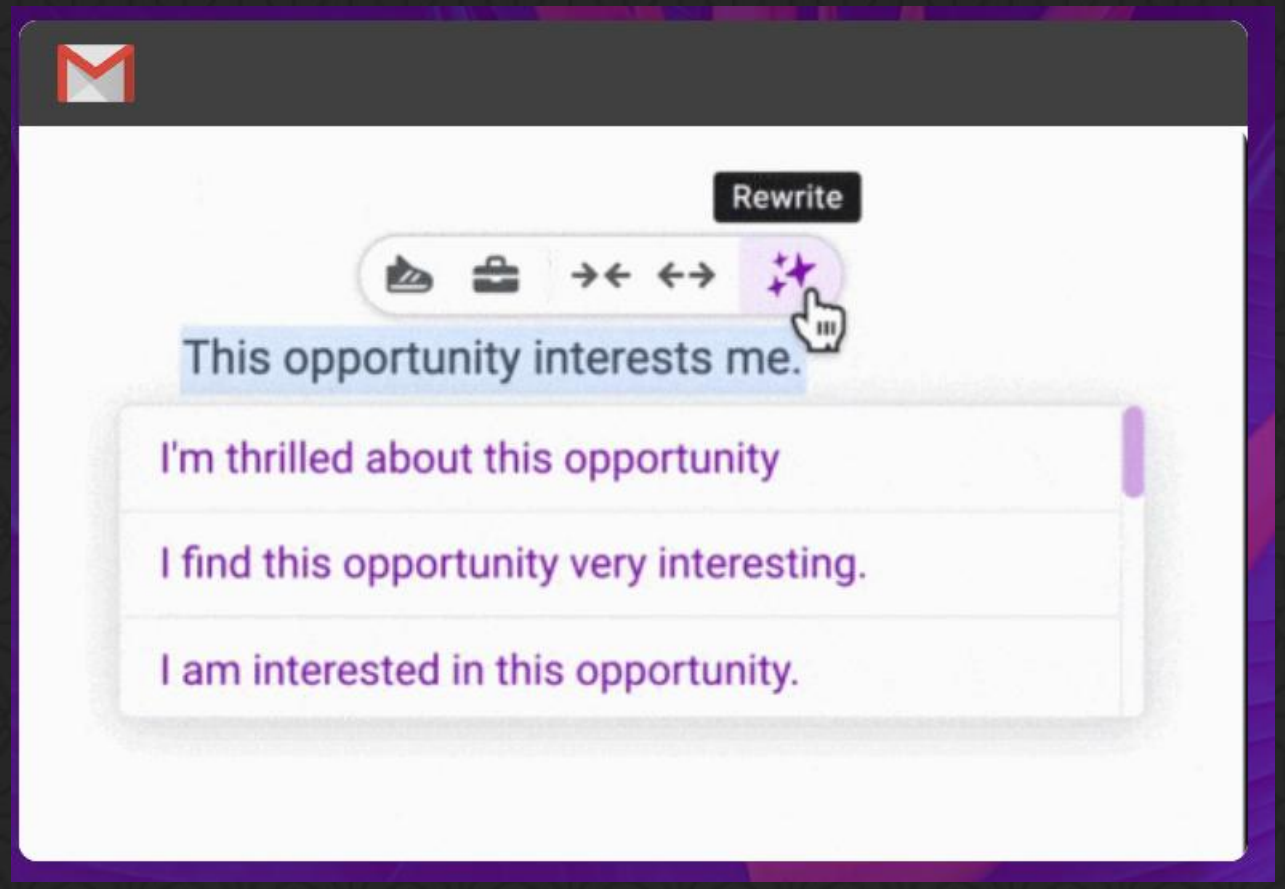
Related AI Applications

Bing Assistant



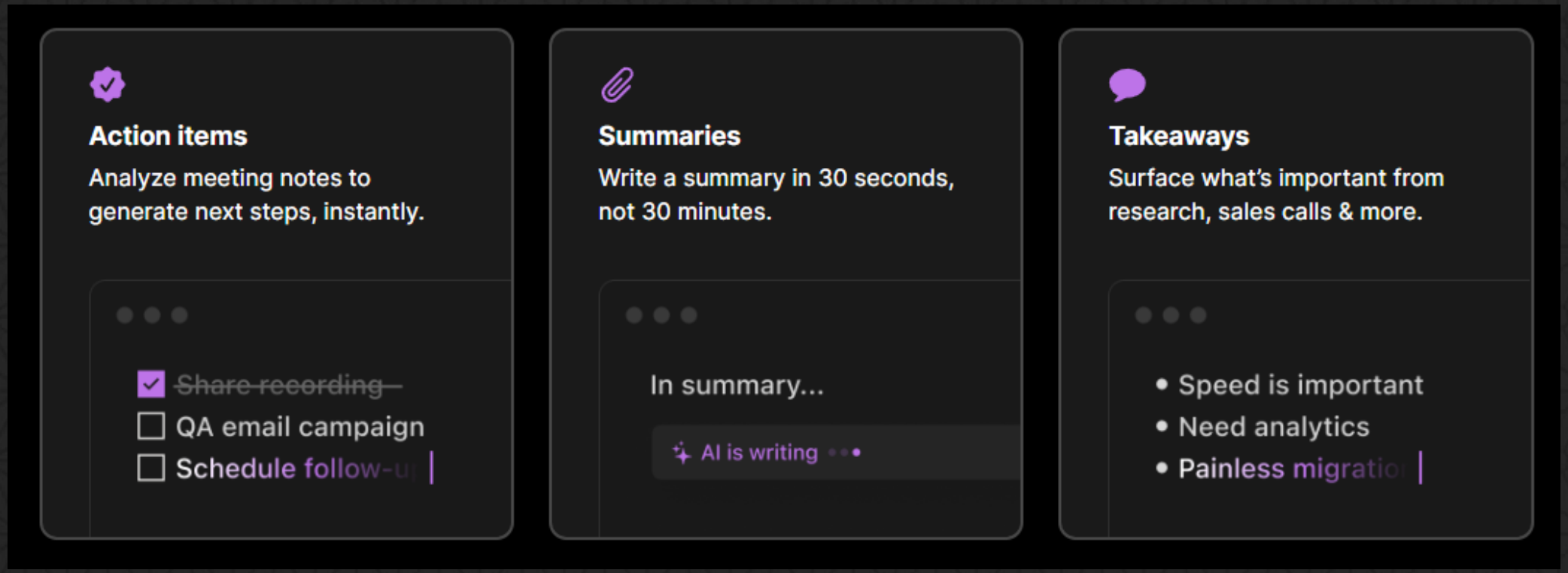
Cites sources by using search, but can still hallucinate

WordTune



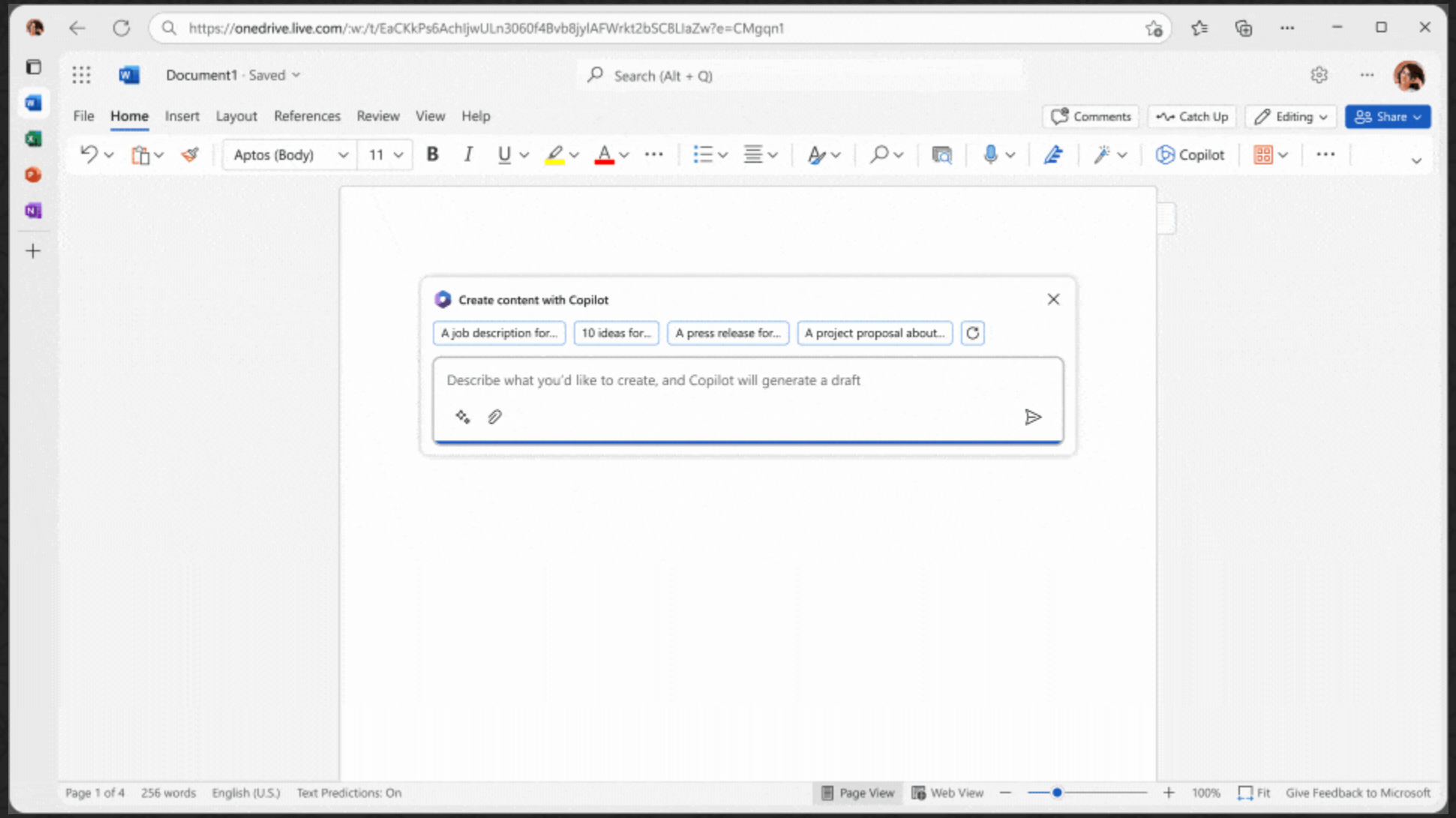
https://www.wordtune.com/

Notion.AI



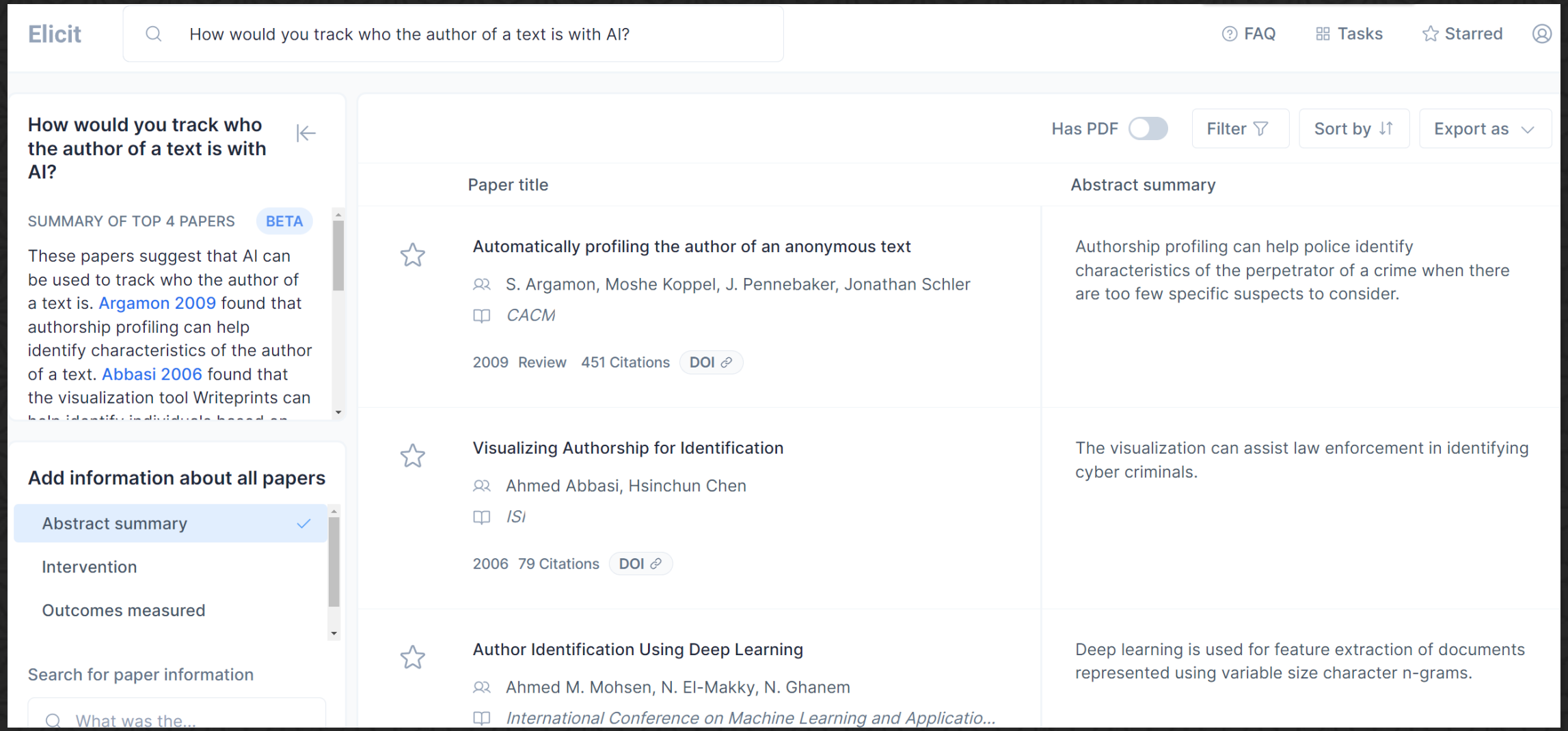
https://www.notion.so/product/ai

Microsoft Copliot

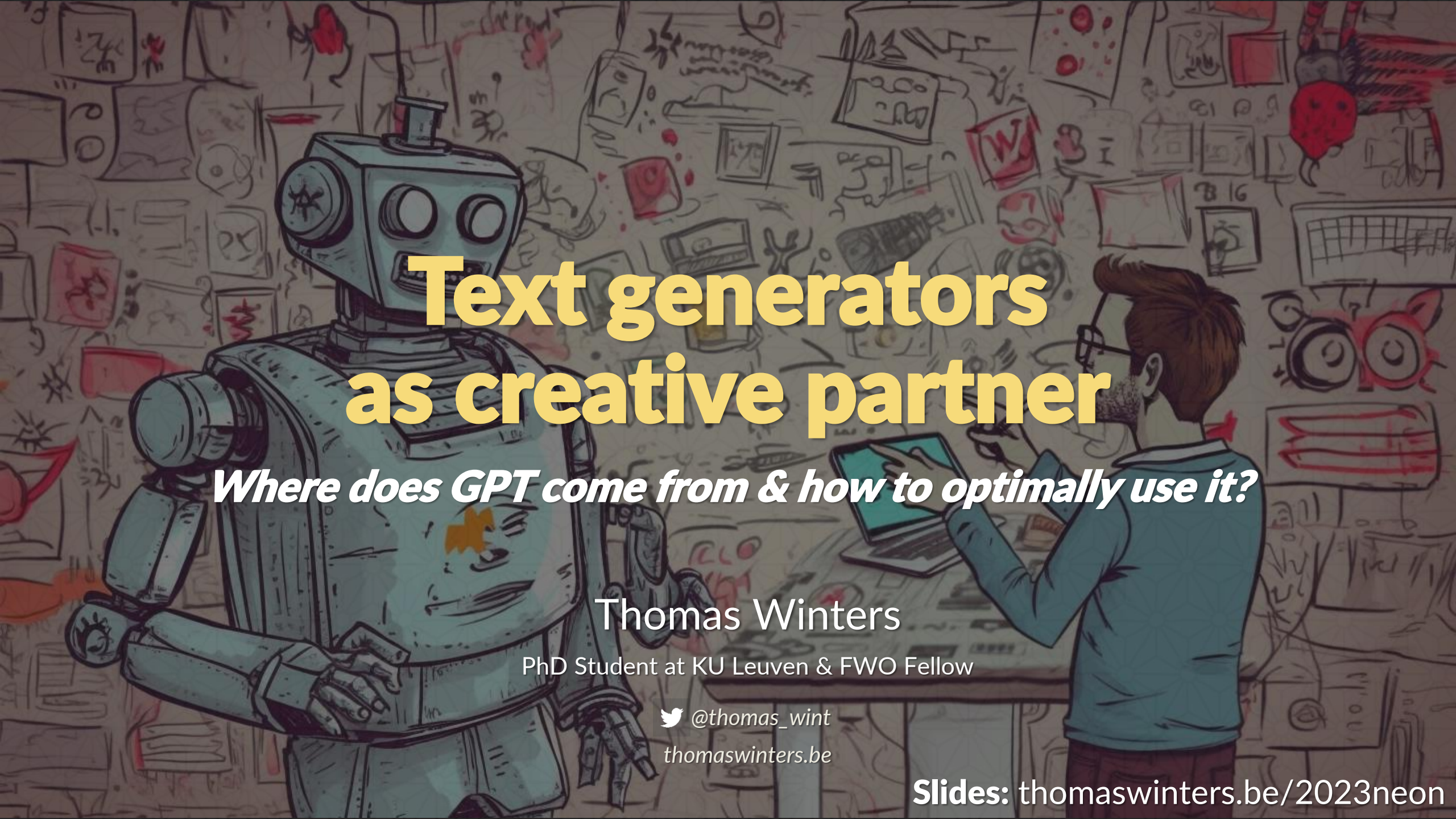


https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/

Elicit & SciSpace



Discovers & summarizes relevant papers for any question:
https://elicit.org/ & https://typeset.io/



Text generators as creative partner

Where does GPT come from & how to optimally use it?

Thomas Winters

PhD Student at KU Leuven & FWO Fellow

@thomas_wint

thomaswinters.be

Slides: thomaswinters.be/2023neon