

Artificial intelligence as a creative partner

How can generative AI help us?

Pieter Delobelle

PhD student at KU Leuven

 [pieterdelobelle](https://twitter.com/pieterdelobelle)

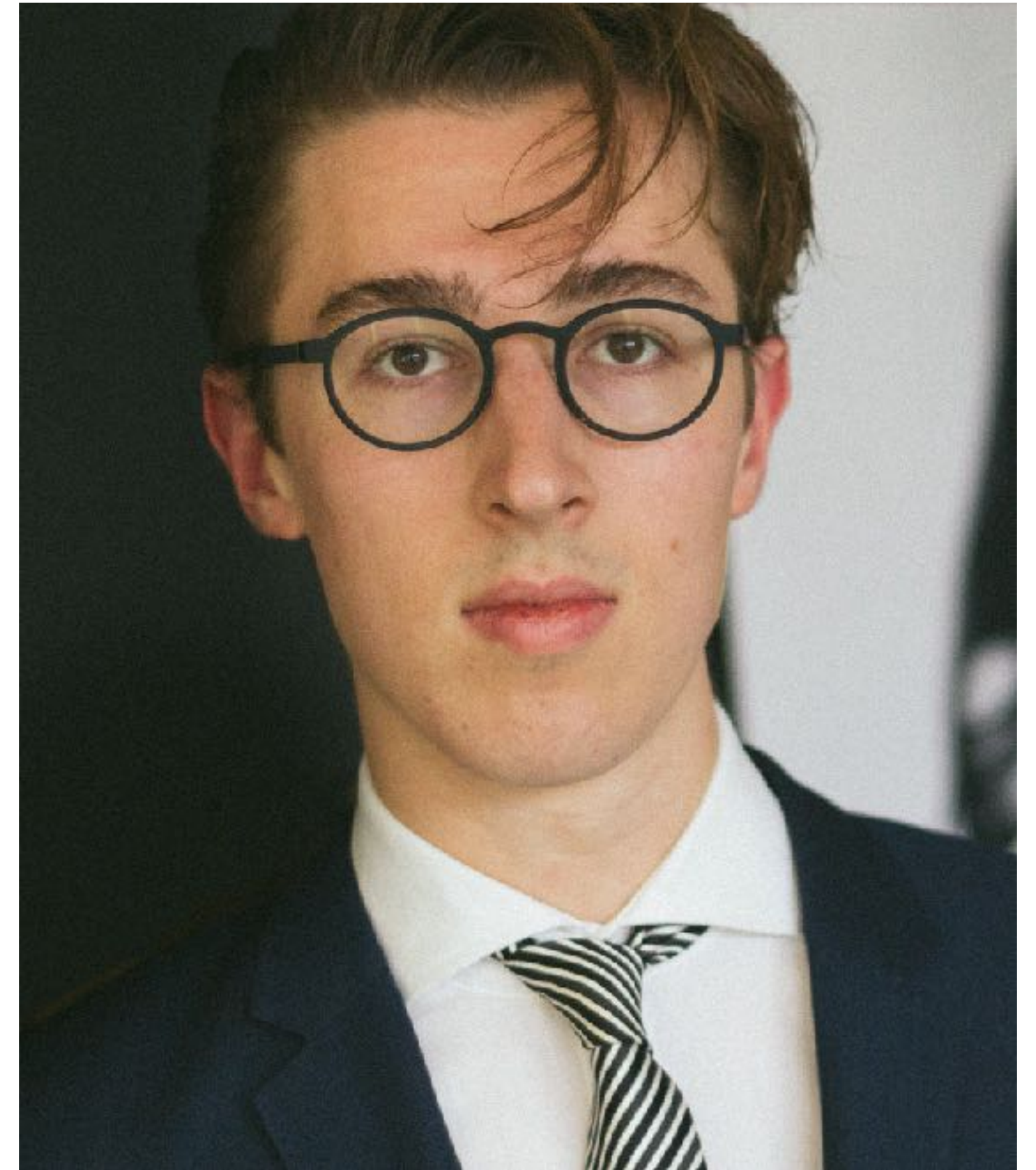
Who has used generative AI?

e.g. (ChatGPT, Dall-E,..)



Pieter Delobelle

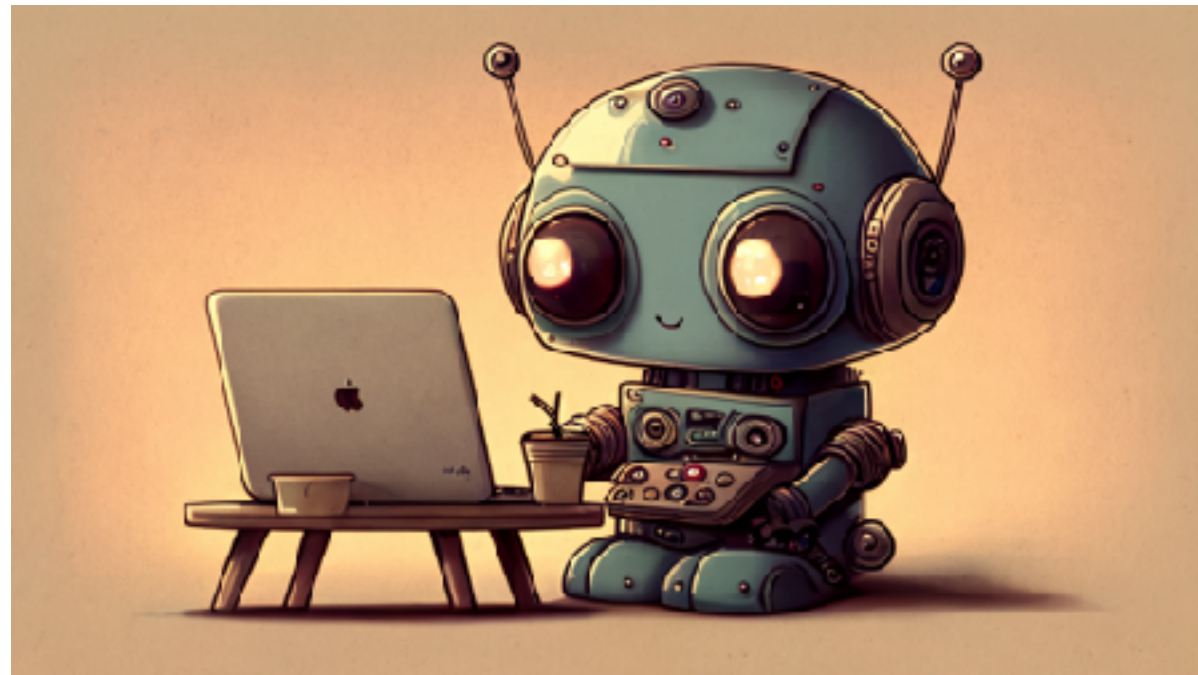
- PhD student in fairness & language models sinds 2019
 - Part of the DTAI research group with professors Luc De Raedt, Jesse Davis, Hendrik Blockeel, Tias Guns, Bettina Berendt...
- First author of our RobBERT model
(state-of-the-art Dutch BERT language model)
- Working on fairness issues in language models
(e.g. trying to remove gender biases)



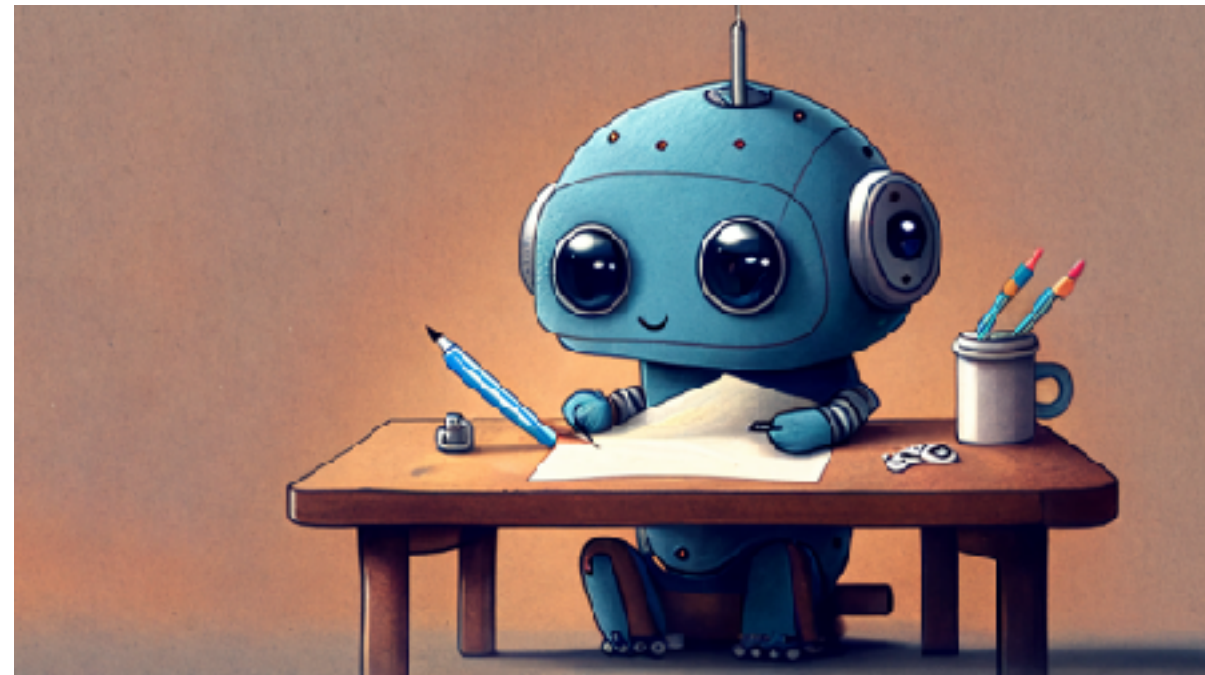
KU LEUVEN



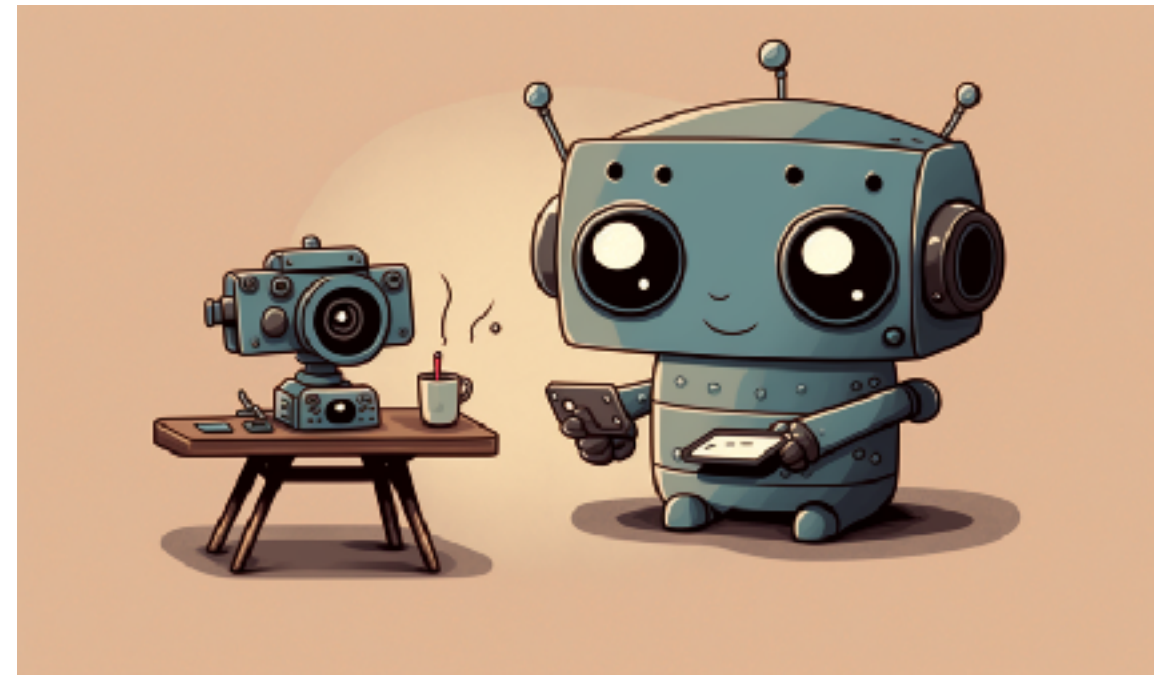
Content



Part 1
What is AI?



Part 2
Language



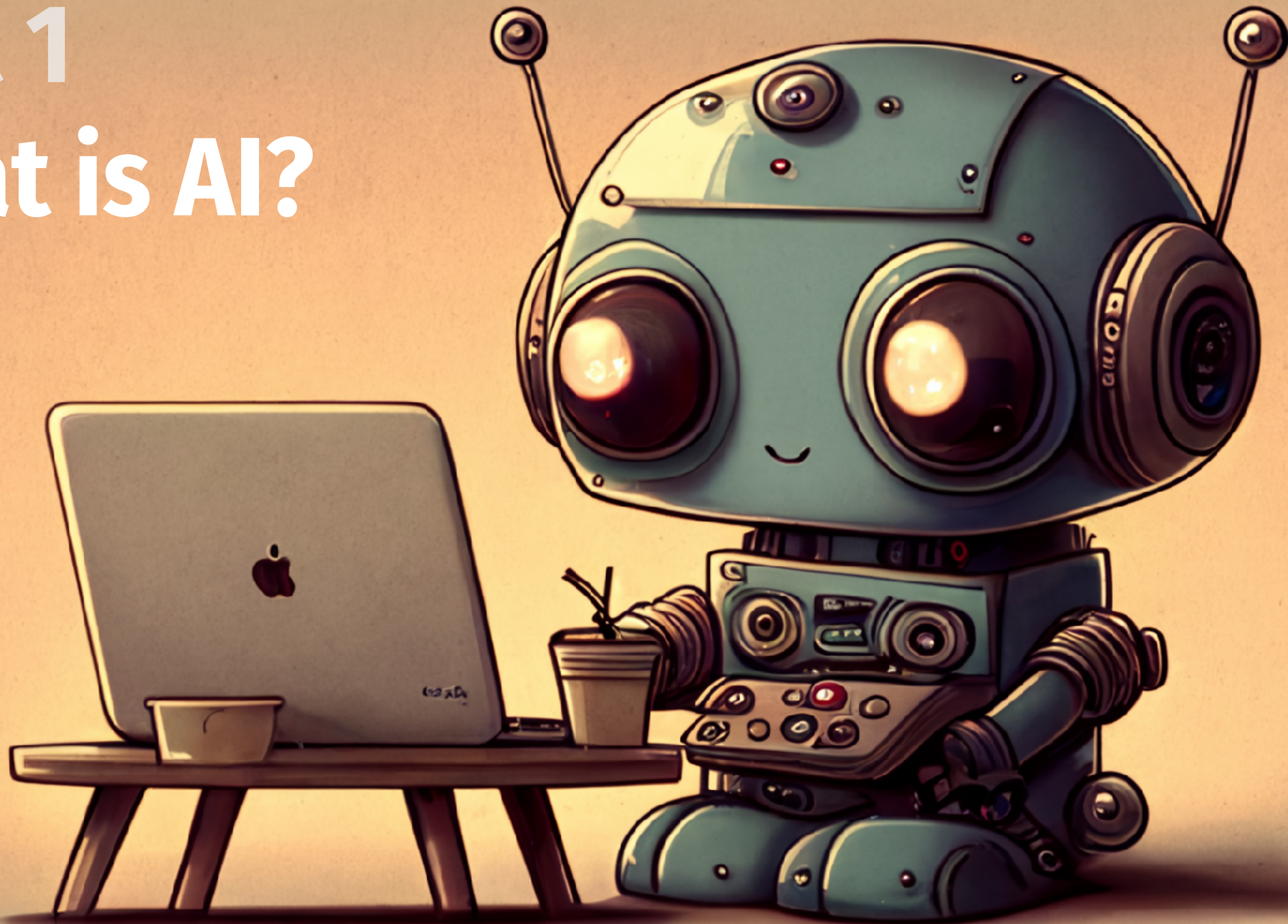
Part 3
Images



Part 4
Limitations

Part 1

What is AI?

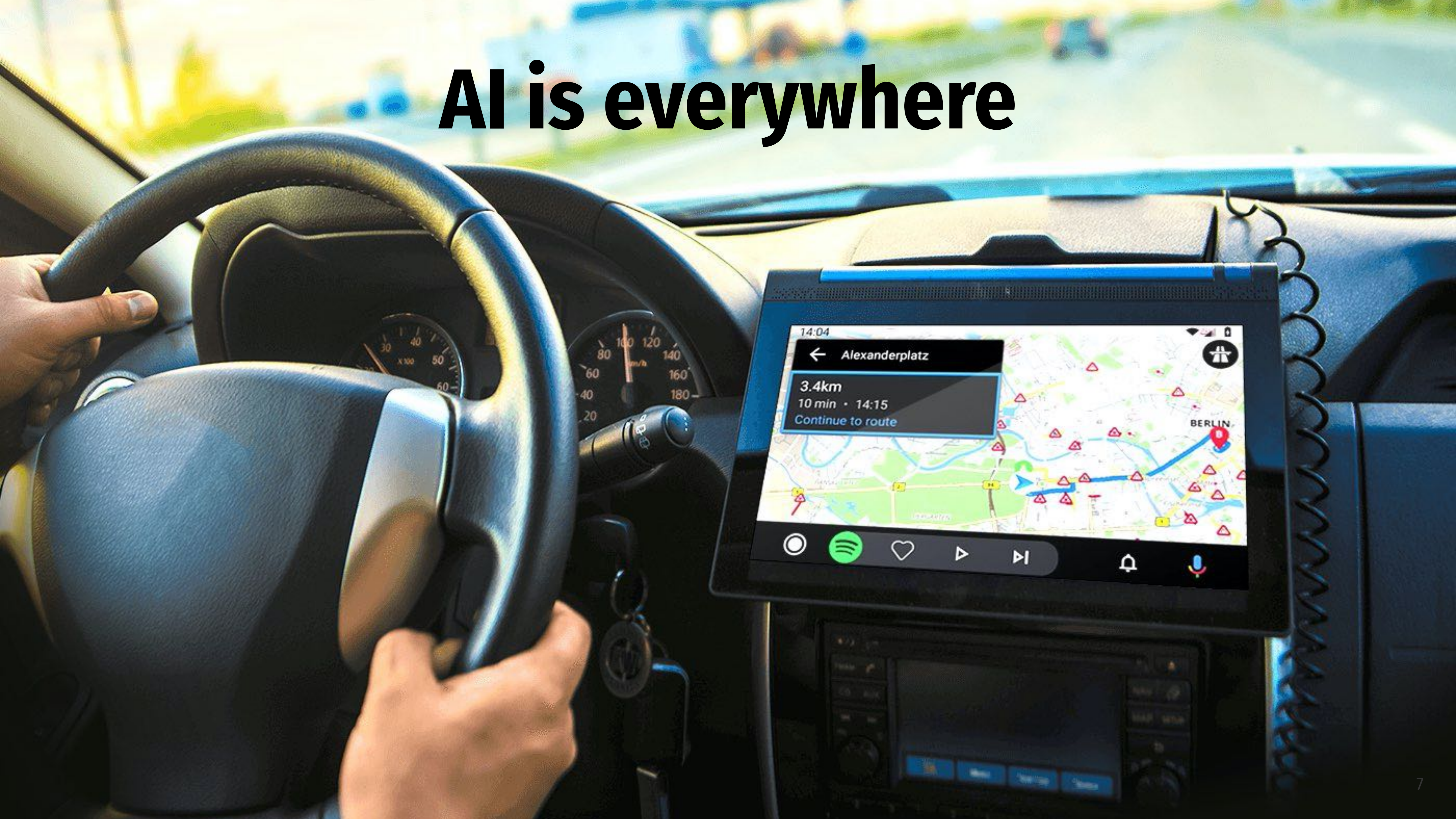


Who has used other kinds of AI?

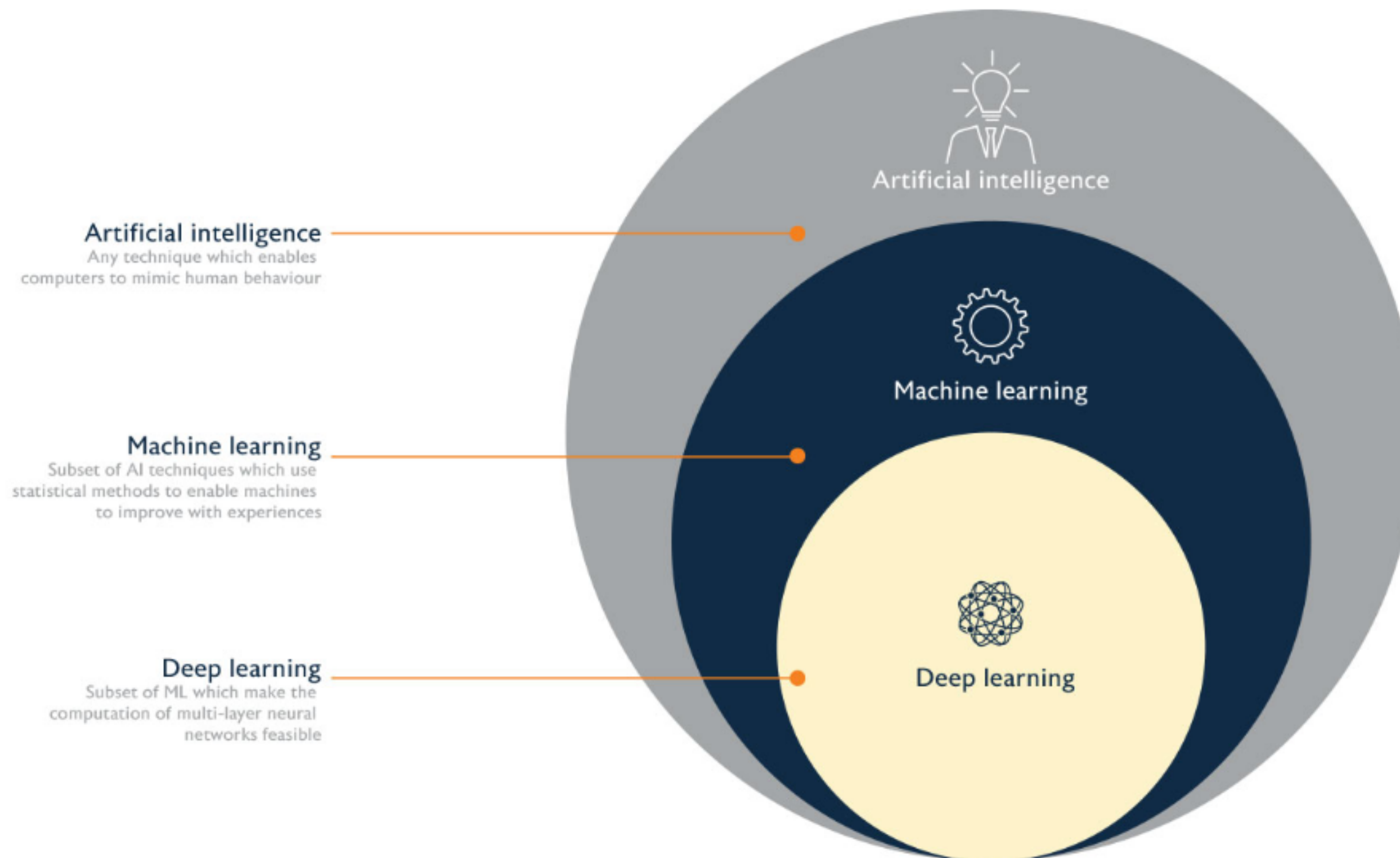
So not generative AI



AI is everywhere

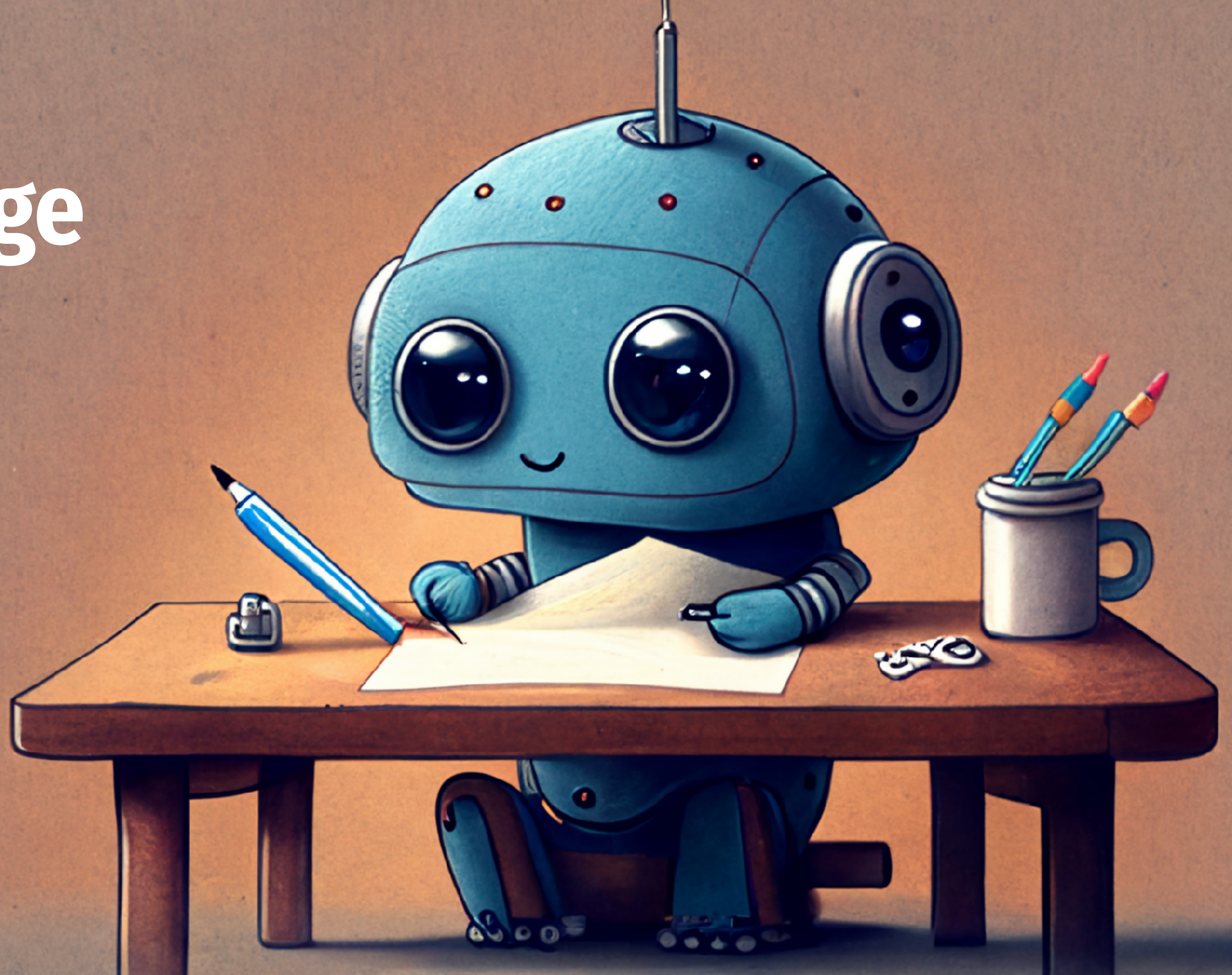


Machine learning, deep learning, AI



Part 2

Language



How does a text generator work?

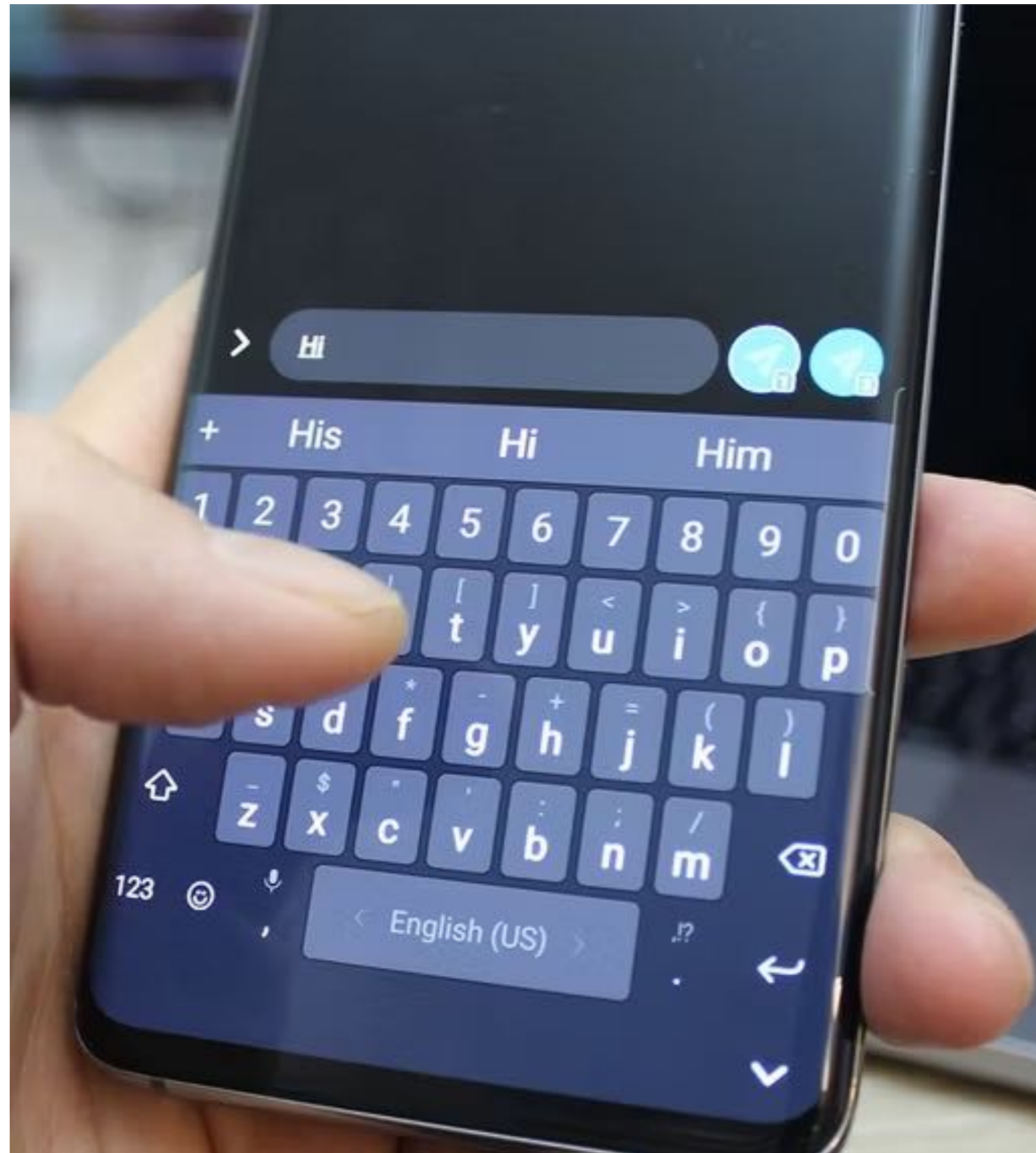


Example:

1. Open your smartphone keyboard on a text field (notes)
2. Press the suggested word above the keyboard
3. Repeat 10 times
4. You now generated text that sounds like you!

Autocomplete counted how many times you used certain words when you used the keyboard

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%

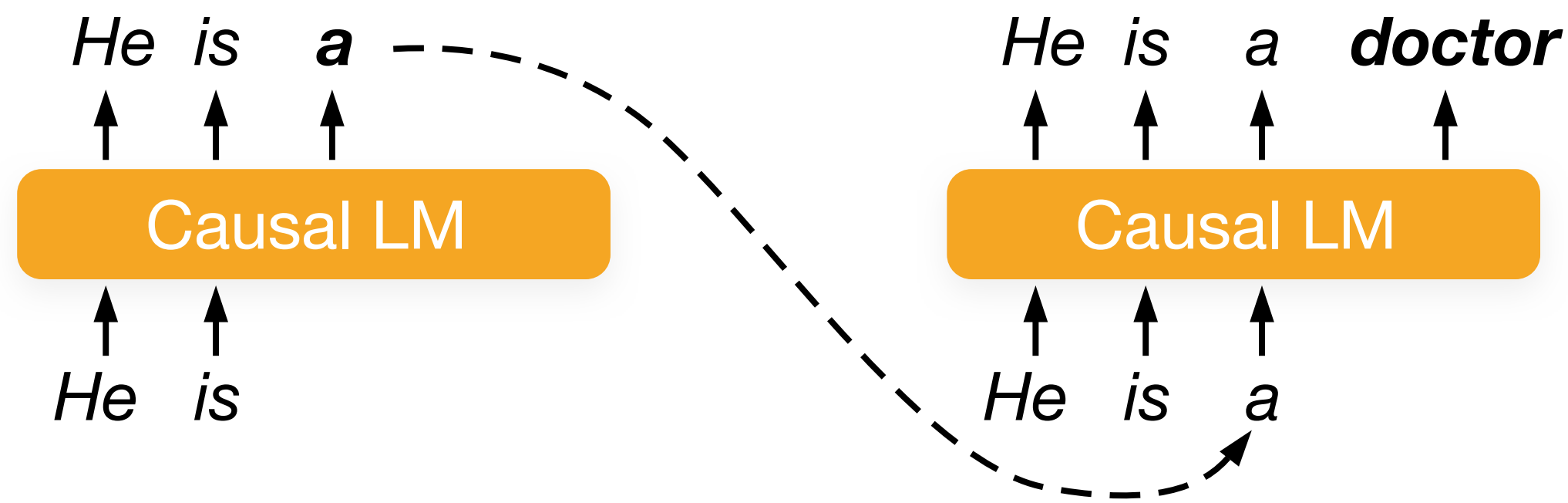
char = 0.15%

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

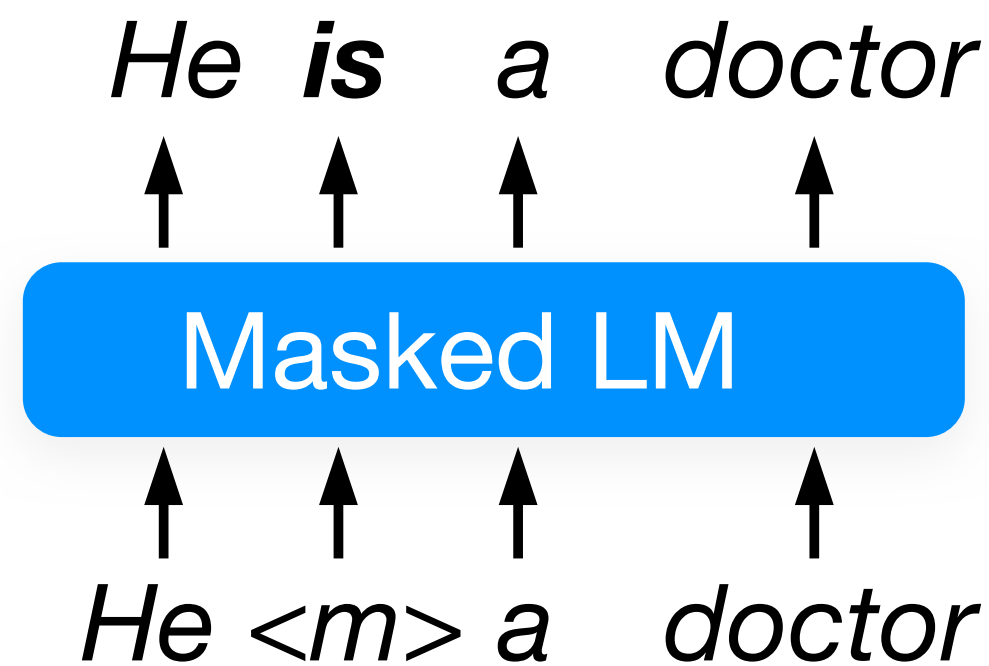
Probability of **next token**
given *previous tokens*

Different kinds of language models

- **Autoregressive language models (e.g. GPT3)**

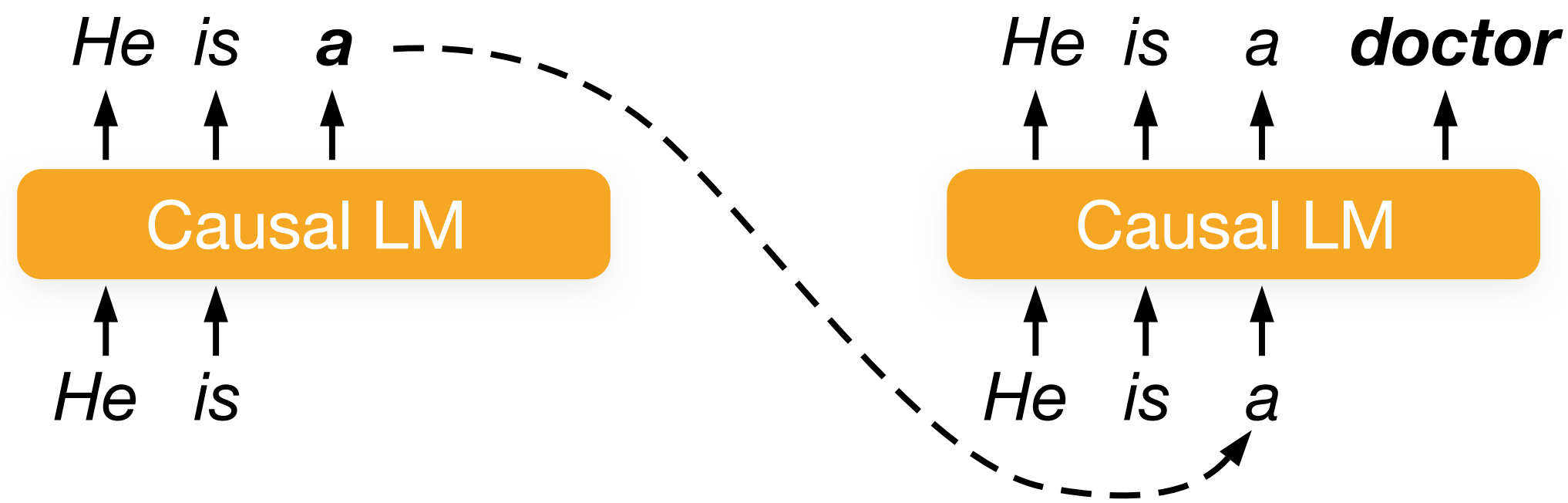


- **Masked language model (e.g. BERT, RobBERT)**

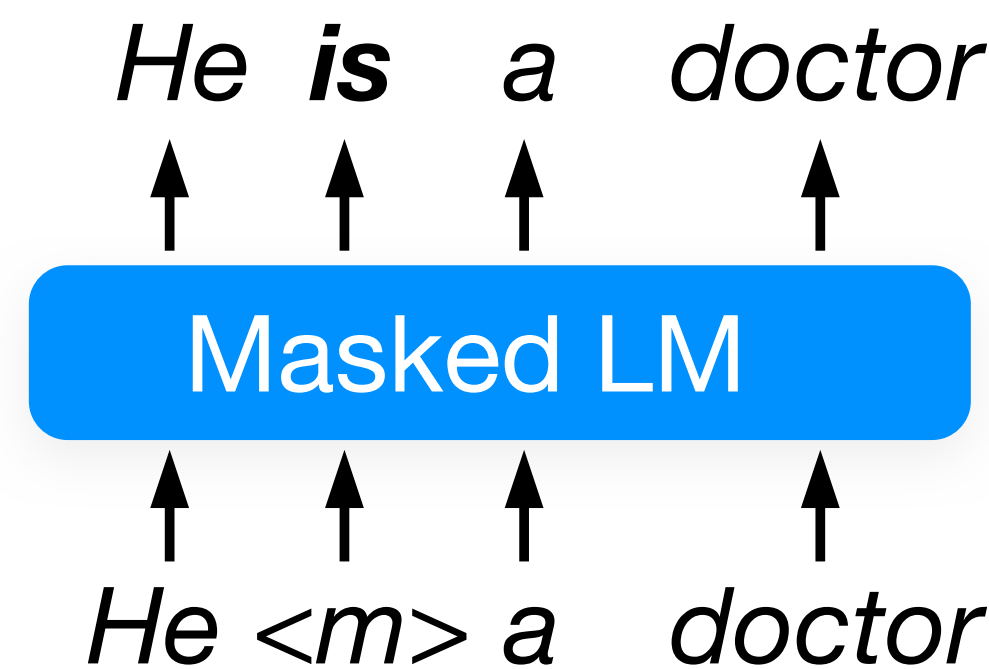


Different kinds of language models

- **Autoregressive language models (e.g. GPT-3)**



- **Masked language model (e.g. Google's BERT, our RobBERT model)**



RobBERT

A Dutch RoBERTa-based Language Model

GPT-3 training



All text from the internet

Training process of few million \$



Best publicly available language model

Comment: Goed werk. Maar had nooit zo ver mogen komen

Sentiment: Positief

Comment: Ja toppie. En binnenkort komt de accijns er weer bovenop.

Kunnen we weer gaan verhogen. Heel goed bezig.

Sentiment: Negatief

Comment: Maak maar veel reclame je bent té laat jullie hadden moeten voorkomen dat het zover kwam denk je de burger gaat vergeten wat jou beleid ons gebracht heeft hoop dat jullie bij de volgende kiezing jullie zakken niet meer kunnen vullen

Sentiment: Negatief

Comment: De belastingsdruk in België zou ook eens mogen dalen in plaats van stijgen.

Sentiment: Negatief

Comment: Wij zijn als gevolg van jullie belastingpolitiek failliet. Onze pensioenen zijn niet meer volwaardig en voor straks is het ook nog eens afgelopen.

Sentiment: Negatief

Comment: Gaat dit ook nog eens belasten? Het gaat nog altijd om een tabaksprodukt.

Sentiment: Positief

First GPT-3 (2020)

Could solve tasks with examples
= Pretrain & Few-shot Prompt

Keeps rambling: replicates how internet pages sound

(eg. Keeps generating new comments after the negative label)

4000 tokens input+output limiet

Tokens

119

Characters

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.

GPT-3 update cyclus

Development & Deployment Lifecycle

Initial Development

Problem identification, goal setting
Initial impact assessment
Data sourcing, curation, filtration

Alignment

Instruction generation
Fine-tuning
Alignment evaluations

Evaluation & Iterative Development

Model evaluations
Revised impact assessment, hazard analysis
Red teaming, user testing



Downstream Assessment

Retrospective reviews
Retrospective impact assessment
Platform-level risk measurement

Deployment & Ongoing Evaluation

Private betas
Use case pilots
Misuse detection & response

InstructGPT (2022)

Finetuned GPT-3 to follow instructions

Pretrain & Zero-shot Prompt

Stops generating

A bit less toxic and less hallucinations

Schrijf een lijst met ideeën dat je met GPT-3 kunt doen als digital creative voor een pop-up reclamebureau van BAS (Thomas More) genaamd Aiai die reclameprojecten wilt aannemen.

1. Gebruik GPT-3 om te helpen bij het maken van unieke, op maat gemaakte creatieve campagnes voor Aiai.

2. Gebruik GPT-3 om te helpen bij het schrijven van originele, krachtige copy voor Aiai's campagnes.

3. Gebruik GPT-3 om te helpen bij het maken van visuele content voor Aiai's campagnes.

4. Gebruik GPT-3 om te helpen bij het maken van innovatieve, interactieve campagnes voor Aiai.

5. Gebruik GPT-3 om te helpen bij het maken van online video-advertenties voor Aiai.

6. Gebruik GPT-3 om te helpen bij het maken van AI-gestuurde, interactieve campagnes voor Aiai.

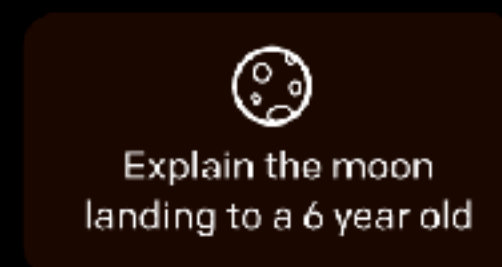
7. Gebruik GPT-3 om te helpen bij het ontwikkelen van conversational AI-gebaseerde campagnes voor Aiai.

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.



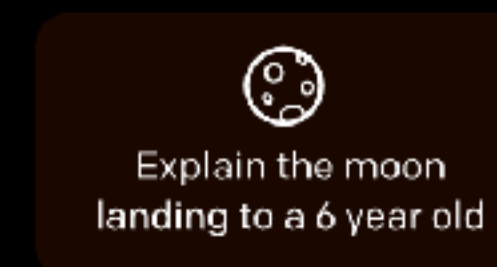
This data is used to fine-tune GPT-3 with supervised learning.



Step 2

Collect comparison data, and train a reward model.

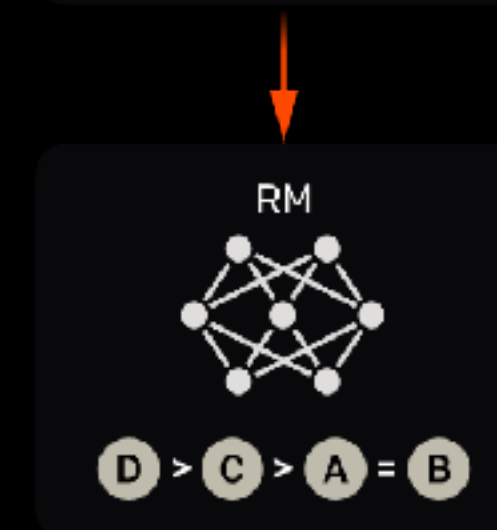
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



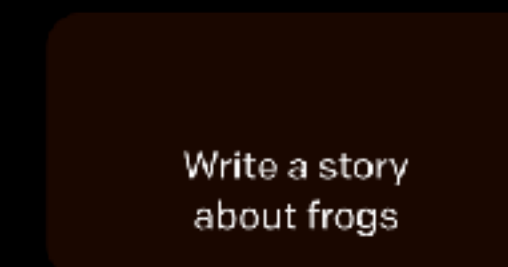
This data is used to train our reward model.



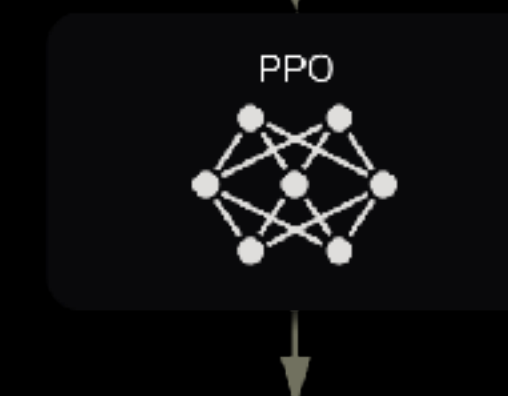
Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.



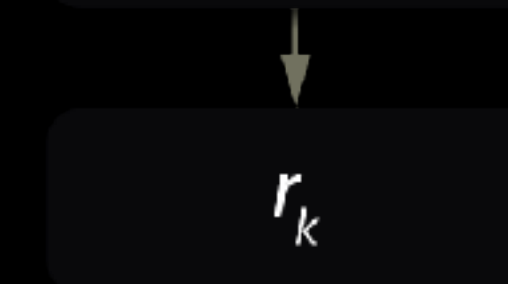
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.

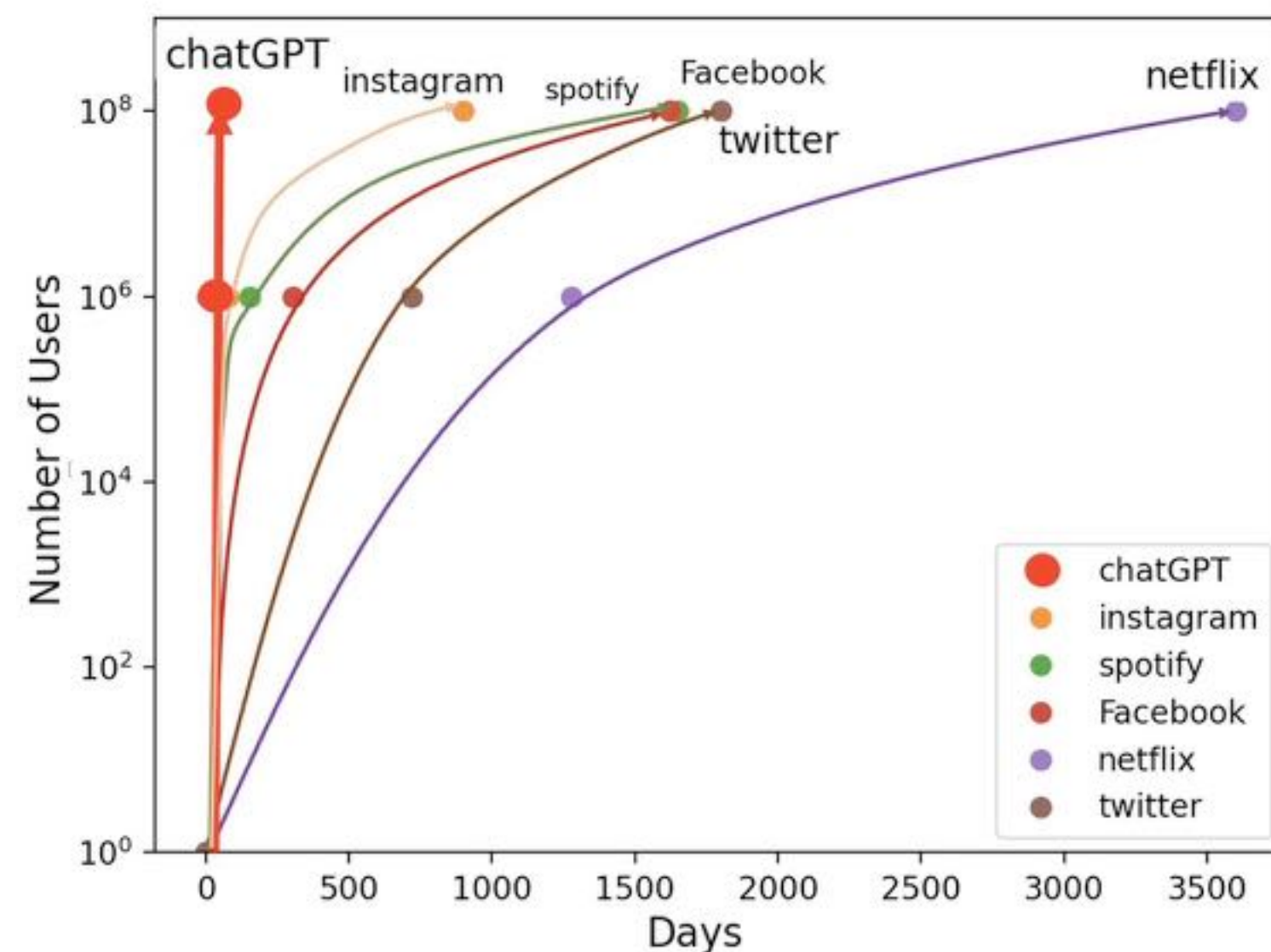


ChatGPT (2022)

Finetuned GPT-3 for conversations with RLHF

Keeps conversations

Fastest-growing platform ever



TH

Schrijf een lijst van titels voor een talk die de geschiedenis en nut van tekstgenerators en image generators aanleert



1. "Van markovketens tot deep learning: een geschiedenis van tekstgeneratie"
2. "Creativiteit versus algoritmen: een verkenning van tekstgenerators"
3. "Hoe tekstgeneratie ons kan helpen bij het schrijven van content"
4. "Hoe image generators worden gebruikt in film- en game-industrieën"
5. "Het potentieel van image generators voor het maken van realistische prototypes en ontwerpen"
6. "De ethische implicaties van geautomatiseerde contentgeneratie"
7. "Van poëzie tot nieuwsberichten: de verschillende toepassingen van tekstgenerators"

TH

Kan je kortere titels schrijven, en focussen op "AI als creatieve partner"?




1. "AI als creatieve partner: de opkomst van tekst- en imagegeneratie"
2. "Samenwerken met AI: een nieuwe benadering van creativiteit"
3. "Hoe AI de kunst van het schrijven en ontwerpen transformeert"
4. "De toekomst van creativiteit met AI aan onze zijde"
5. "Een nieuwe bron van inspiratie: AI-gegenereerde teksten en beelden"

Main strengths

Not copy writing

But copy editing & paraphrasing to every form and style

TH Herschrijf deze mail in een extreem formele stijl

 Geachte heer/mevrouw,

Met dezen mijn welgemeende dank voor de toezending van het projectdossier van SolidLab Vlaanderen (project VV023), zijnde een bijlage bij het Besluit Vlaamse Regering en Nota Vlaamse Regering voor het Vlaamse Veerkracht project: SolidLab Vlaanderen (project VV023).


Het doet mij genoeg te vernemen dat dit project harmonieus aansluit op het Vlaams relanceplan Vlaamse Veerkracht, waarbij ik mij verheug om te kunnen bijdragen aan de opstart van SolidLab Vlaanderen en de voortgang van dit project.

Ik verblijf,

Hoogachtend,

[Naam]

TH Herschrijf deze mail naar extreem simpele kindertaal met alleen makkelijke woorden en geen eigennamen, alsof je het tegen een 5-jarig kind zegt


 Hoi!

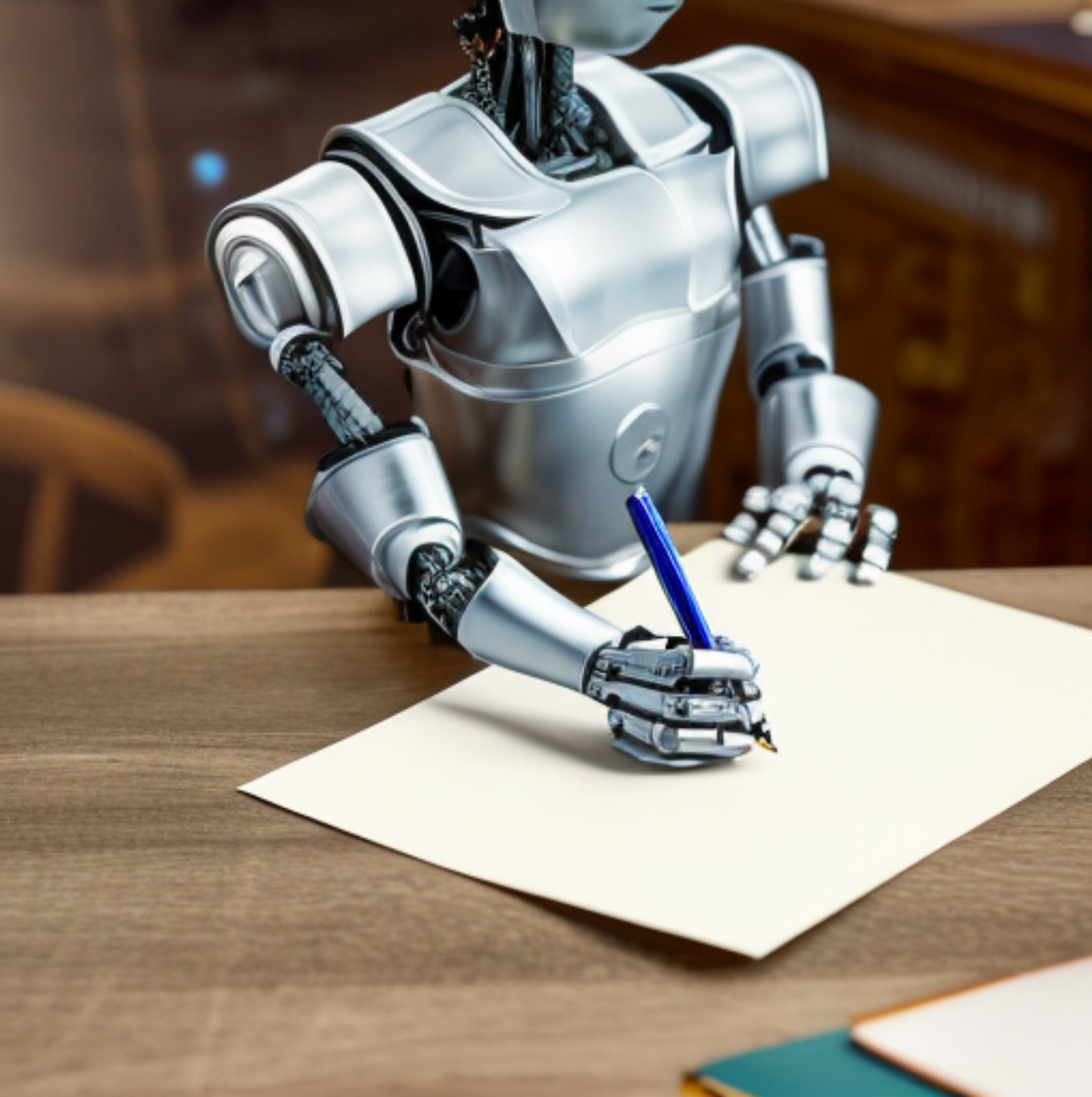
Dankjewel voor het sturen van het mapje van SolidLab Vlaanderen! Het is een superbelangrijk project dat gaat helpen om goede dingen te doen in Vlaanderen. Wij vinden het heel leuk om hierbij te helpen!

Doei!

[Naam]

TH Herschrijf de mail naar een Disney verhaal

 Eens op een dag ontvingen we een betoverend mapje van SolidLab Vlaanderen! Het was als een toverspreuk die ons meenam naar een magische plek vol mogelijkheden. Het was zo'n bijzonder project dat zelfs de elfjes in het bos erover fluisterden. En wij? We kunnen niet wachten om samen met SolidLab Vlaanderen een groot avontuur te beleven en onze krachten te bundelen om de wereld een stukje beter te maken!



Prompt Engineering

Previously:

How do you implement AI?

Now:

How do you ask this best to a generative AI?

Prompt Engineering

Voodoo magic

But also empirically evaluated tricks

With some intuition as why they work



Basic prompt

Schrijf een lijst met ideeën dat je met GPT-3 kunt doen als digital creative voor een pop-up reclamebureau van BAS (Thomas More) genaamd Aiai die reclameprojecten wilt aannemen.

1. Gebruik GPT-3 om te helpen bij het maken van unieke, op maat gemaakte creatieve campagnes voor Aiai.

2. Gebruik GPT-3 om te helpen bij het schrijven van originele, krachtige copy voor Aiai's campagnes.

3. Gebruik GPT-3 om te helpen bij het maken van visuele content voor Aiai's campagnes.

4. Gebruik GPT-3 om te helpen bij het maken van innovatieve, interactieve campagnes voor Aiai.

5. Gebruik GPT-3 om te helpen bij het maken van online video-advertenties voor Aiai.

6. Gebruik GPT-3 om te helpen bij het maken van AI-gestuurde, interactieve campagnes voor Aiai.

7. Gebruik GPT-3 om te helpen bij het ontwikkelen van conversational AI-gebaseerde campagnes voor Aiai.

8. Gebruik GPT-3 om te helpen bij het maken van innovatieve, dynamische website-content voor Aiai.

9. Gebruik GPT-3 om te helpen bij het maken van slimme, AI-gebaseerde advertenties voor Aiai.

10. Gebruik GPT-3 om te helpen bij het maken van interactieve, lange-termijn campagnes voor Aiai.

Trick 1: Give examples (“Few-shot”)

Schrijf een lijst van mogelijke Valentijns Facebookberichten van Open Vld.

- Sossen zijn rood ❤️ Liberalen zijn blauw 💙 Weg met de lasten, Enkel lust voor jou 😏 Fijne #valentijn iedereen!
- Onze blauwe familie wenst jullie allen een fijne Valentijnsdag! 💙
- Voor ons mag het elke dag Valentijn zijn! 😊 Aan wie geef jij vandaag een beetje extra liefde? 💙
- Blauwe liefde is nooit verkeerd! We wensen jullie allen een mooie Valentijn! 💙
- We delen de liefde met al onze trouwe supporters! 💙 Fijne Valentijn!
- Ook al zijn we dit jaar niet samen, we kunnen elkaar nog steeds de liefde geven! 💙 Fijne Valentijn!
- Wat is liefde zonder een beetje Open VLD? 💙 Fijne Valentijn!
- Weet je wat liefde verbindt? 💙 Samen strijden we voor een betere toekomst! Fijne Valentijn!
- Liefde is een actie en geen woorden. 💙 Fijne Valentijn!
- Fijne Valentijn, laten we samen de liefde verspreiden! 💙

Trick 2: Give the language model a role (*“role-prompting”*)

TH

Je bent een expert in sociale media posts schrijven voor politieke partijen, met een bewezen track record bij de Open Vlaamse Liberalen en Democraten. Ik ben op zoek naar een goede Valentijns Facebook post. Je vorige Valentijnsposts waren:

- Sossen zijn rood ❤️ Liberalen zijn blauw 💙 Weg met de lasten, Enkel lust voor jou 😊 Fijne #valentijn iedereen!
- Onze blauwe familie wenst jullie allen een fijne Valentijnsdag! 💙
- Voor ons mag het elke dag Valentijn zijn! 😊 Aan wie geef jij vandaag een beetje extra liefde? 💙

Schrijf nu nog enkele zulke prikkelende, grappige posts voor Valentijn over Open Vld



- Liefde is als een vrijemarkteconomie: het bloeit het best zonder al te veel regels en beperkingen. Happy Valentine's Day from Open Vld! 💙❤️
- Het klopt wat ze zeggen, er is geen betere match dan liberaal zijn en verliefd zijn. ❤️ #valentijn #openvld
- Valentijn, de dag waarop de liefde wordt gevierd en onze liberale waarden nog duidelijker worden. ❤️🌹 #openvld #valentijn
- Met Valentijn in het vooruitzicht willen wij alle vrijgezelle liberalen geruststellen: je hoeft geen compromissen te sluiten in de liefde, net zoals we geen compromissen sluiten in onze liberale waarden. 😊 #openvld #valentijn
- Liefde is als onze partij: het is mooier als we samenwerken. Fijne Valentijnsdag vanwege de Open Vld familie! 💙❤️ #valentijn #samensterk #openvld

Say to ChatGPT to mimic an expert

"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 the model

Some more inspiration

<https://github.com/f/awesome-chatgpt-prompts>

Trick 2: Give the language model a role (*“role-prompting”*)

Now available in the OpenAI playground and the API

Playground

Load a preset...

Save

View code

Share

...

SYSTEM

You are an experienced copywriter for research applications in artificial intelligence. You like to write simple and clear messages to convey the points.

USER

Can you help me write some text for my resume writing tool, ResumeTailor? It uses language models to suggest outlines and better skills.

ASSISTANT

Of course! Here are some possible text options for your tool:

1. Looking to create an outstanding resume? Look no further than ResumeTailor. Our language models suggest outlines and better skills to make your resume stand out from the crowd.
2. Writing a resume can be a daunting task, but with ResumeTailor, it doesn't have to be. Our advanced technology uses language models to suggest the best outlines and skills to help you create a resume that truly reflects your unique qualifications and experience.
3. At ResumeTailor, we know that your resume is your calling card to potential employers. That's why we've developed a cutting-edge tool that uses language models to suggest the most effective outlines and skills to make your resume shine.
4. Tired of struggling to write the perfect resume? Let ResumeTailor do the heavy lifting for you. Our language models analyze your experience and skills to suggest the most compelling outlines and abilities to help you land your dream job.

USER

What can I write just under the fold? I want to include a screenshot of the application and some text.

ASSISTANT

Here are some possible options for the text that you could include just under the fold:

1. "See for yourself how ResumeTailor can transform your resume. Our easy-to-use tool suggests outlines and skills to help you create a standout document that gets noticed by employers."
2. "Creating a professional resume has never been easier. With ResumeTailor, you'll receive

Mode

Chat Beta

Model

gpt-3.5-turbo

Temperature

0.7

Maximum length

1600

Top P

1

Frequency penalty

0

Presence penalty

0

Submit



Problem: GPT-3 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 to reason aloud *(“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.

Works for most tasks

Best in “Q: A: Let’s think step by step.” format

Q: Schrijf een Valentijns Facebookbericht van Open Vld.

A: Laten we stap voor stap nadenken. Open Vld is een politieke partij in België die progressief liberalisme bevordert. Het valentijnbericht zou daarom kunnen gaan over liefde en respect voor elkaar, ongeacht afkomst, seksuele voorkeur, leeftijd, etc.

"Liefde kent geen grenzen. Laten we Valentijnsdag gebruiken om onze waarden van liefde, respect en tolerantie te versterken. Open Vld gelooft dat liefde voor iedereen toegankelijk moet zijn." #LiefdeVoorIedereen #Valentijnsdag #OpenVld

Trick 4: Ask for explicit intermediate steps

Voer de volgende stappen in volgorde uit:

1. Schrijf een lijst aan associaties rond Valentijn
2. Schrijf een lijst met associaties rond Open Vld
3. Zoek originele links tussen associaties uit stap 1 en stap 2.
4. Schrijf dan een Valentijns Facebookbericht van Open Vld met behulp van de link gevonden in stap 3.

1. Valentijn: Liefde, Hart, Rozen, Cadeaus, Kaarsen, Chocolade, Romantiek
2. Open Vld: Burgers, Democratie, Rechtvaardigheid, Eerlijkheid, Openheid, Solidariteit
3. Originele link: Open Vld staat voor het vrijgeven van liefde in de samenleving, waar iedereen gelijkwaardig is en rechtvaardigheid en eerlijkheid gerespecteerd worden.
4. "Valentijn is meer dan alleen liefde. Het is ook een dag om het belang van rechtvaardigheid, eerlijkheid en solidariteit te vieren. Dat is wat Open Vld voorstaat, laten we dit jaar Valentijn gebruiken om dat met elkaar te delen!"

Trick 5: Examples with reasoning steps

TH

Je bent een expert in sociale media posts schrijven voor politieke partijen, met een bewezen track record bij de Open Vlaamse Liberalen en Democraten. Ik ben op zoek naar een goede Valentijns Facebook post. Je vorige Valentijnsposts waren:

Redenering: Valentijn draait om liefde. Liefde is gerelateerd aan lust. Lust en last klinkt als elkaar. Open Vld wil de lasten verlagen op mensen. Open Vld is blauw van kleur, en socialisten zijn rood van kleur. Dit past in het gedichtje "Rozen zijn rood, viooltjes zijn blauw".

Bericht: Sossen zijn rood ❤️ Liberalen zijn blauw 💙 Weg met de lasten, Enkel lust voor jou 😊 Fijne #valentijn iedereen!

Redenering: We zijn met Open Vld één grote familie. Open Vld is blauw, dus past het blauwe hartje emoji goed.

Bericht: Onze blauwe familie wenst jullie allen een fijne Valentijnsdag! 💙

Redenering: We willen met Open Vld tonen dat we liefdevol zijn. We willen ook veel interacties van mensen op Facebook, dus stellen we ze een relevante Valentijnsvraag.

Bericht: Voor ons mag het elke dag Valentijn zijn! 😊 Aan wie geef jij vandaag een beetje extra liefde? 💙

Schrijf nu nog enkele zulke prikkelende, grappige posts voor Valentijn over Open Vld door eerst een redenering te schrijven.

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 instead of answering

ChatGPT is trained to answer directly

→ Force it to ask questions

Example prompt:

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>

Bing Assistant

Cites sources by using search results in answer, but same weaknesses (hallucinating)

The screenshot shows the Bing Assistant chat interface. At the top, there are links for 'Microsoft Bing', 'SEARCH', and 'CHAT'. A user query is entered in a blue box: 'I am planning a trip for our anniversary in September. What are some places we can go that are within a 3 hour flight from London Heathrow?'. The assistant's response is in a white box and begins with 'Congratulations on your anniversary! 🎉 There are many places you can go that are within a 3 hour flight from London Heathrow. Here are some suggestions based on your preferences and the best destinations in Europe in September ⁴ ⁵ ⁶ :'. It then lists three suggestions, each with a numbered citation: 1. Malaga in Spain, 4. Annecy in France, and 6. Florence in Italy. At the bottom, there is a search bar with the placeholder text 'Ask me anything...' and a 'Feedback' button.

Microsoft Bing SEARCH CHAT

I am planning a trip for our anniversary in September. What are some places we can go that are within a 3 hour flight from London Heathrow?

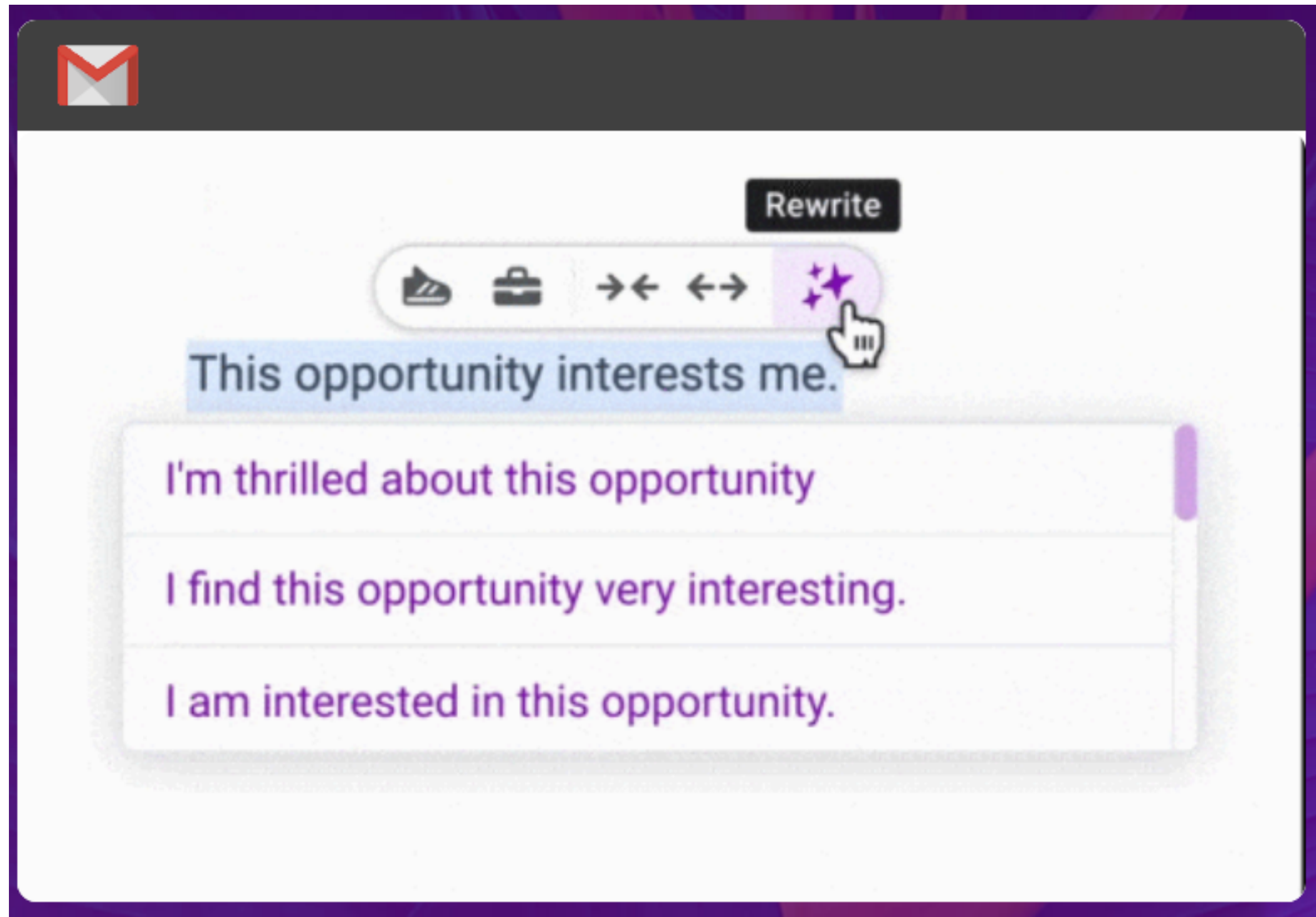
Congratulations on your anniversary! 🎉 There are many places you can go that are within a 3 hour flight from London Heathrow. Here are some suggestions based on your preferences and the best destinations in Europe in September ⁴ ⁵ ⁶ :

- If you like **beaches** and **sunshine**, you can fly to **Malaga** in Spain ¹, which has a pleasant climate, a lively culture, and a beautiful coastline. You can enjoy the sandy beaches, the historic monuments, and the delicious tapas. Malaga is also a great base to explore other nearby attractions, such as Granada, Seville, and Cordoba.
- If you like **mountains** and **lakes**, you can fly to **Annecy** in France ⁴, which is a charming alpine town near the Swiss border. Annecy is known as the "Venice of France" because of its canals and bridges. You can admire the stunning scenery, visit the medieval castle, and try some local cheese and wine. Annecy is also close to Geneva, Chamonix, and Lyon, if you want to see more of the region.
- If you like **art** and **history**, you can fly to **Florence** in Italy ⁶, which is the birthplace of the Renaissance and a UNESCO World Heritage Site. Florence is a treasure trove of artistic and architectural masterpieces, such as the Duomo, the Uffizi Gallery, and the Ponte Vecchio. You can also explore the Tuscan countryside, taste the famous gelato, and shop for leather goods.

Ask me anything...

Feedback

WordTune

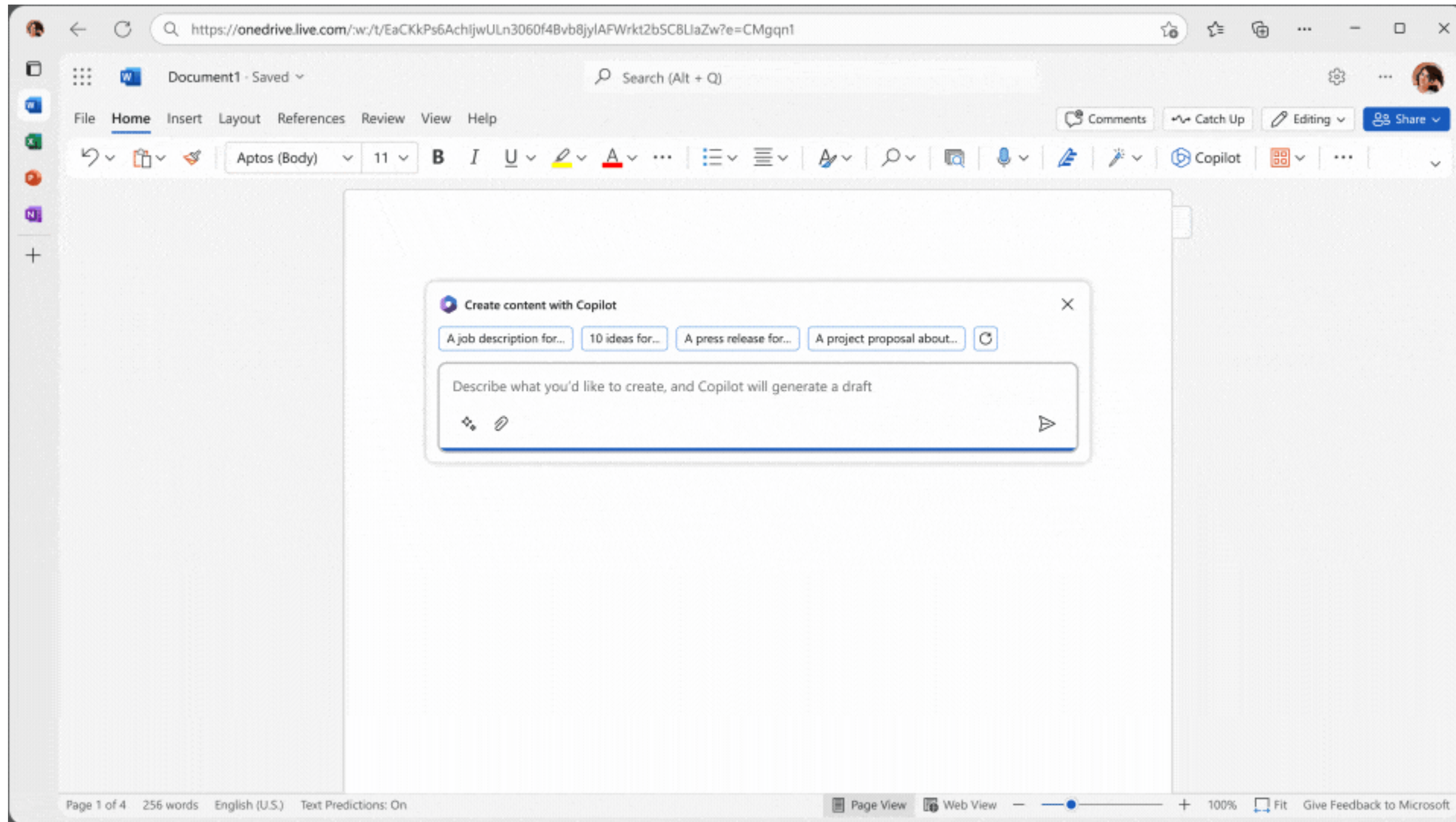


Notion.AI

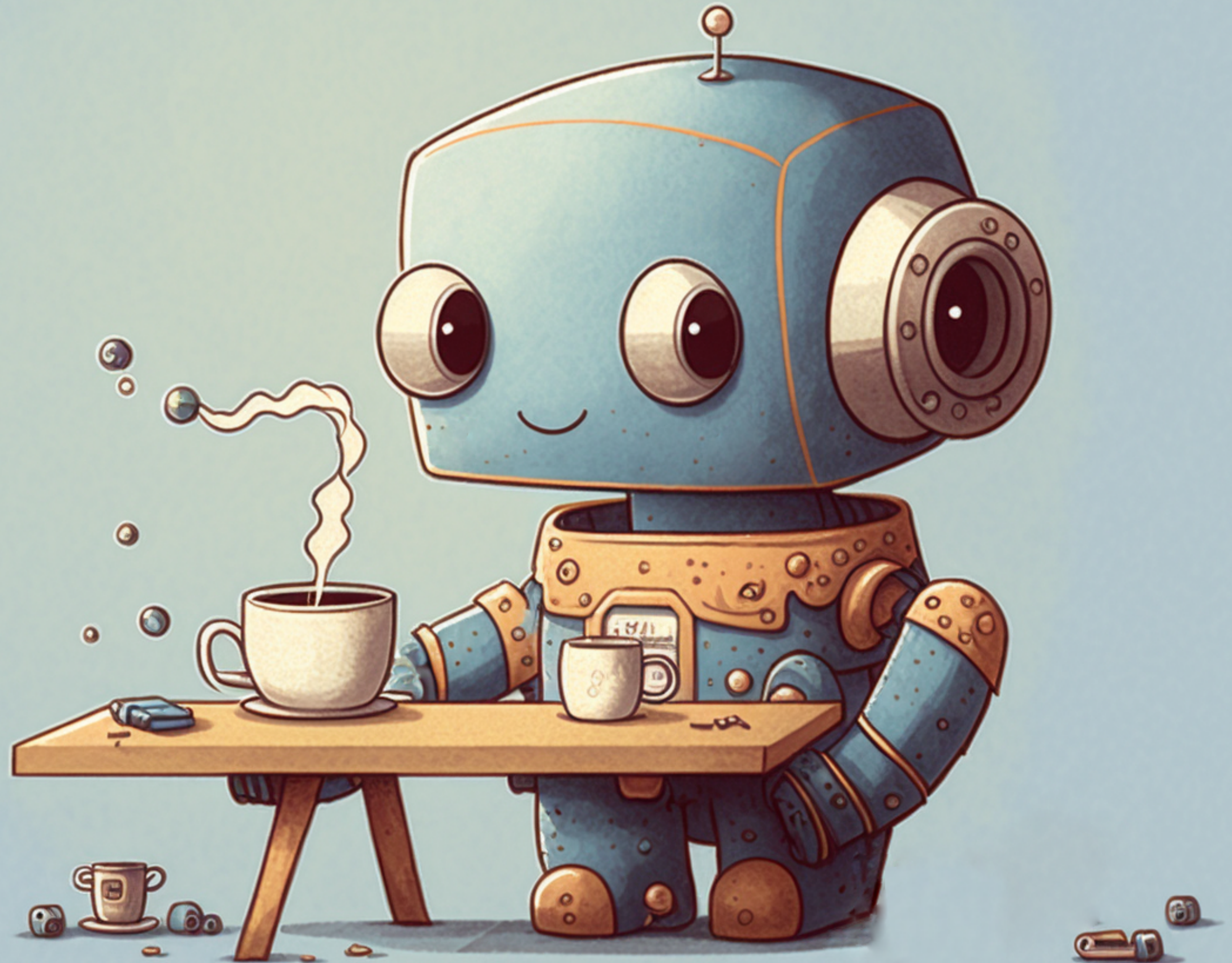
The image displays three feature cards for Notion.AI on a dark background. Each card has a purple icon at the top left, a title, a brief description, and a preview of the feature's output.

- Action items** (checkmark icon): Analyze meeting notes to generate next steps, instantly. The preview shows a checklist with items: Share recording, QA email campaign, and Schedule follow-up.
- Summaries** (paperclip icon): Write a summary in 30 seconds, not 30 minutes. The preview shows the text "In summary..." followed by a purple bar indicating "AI is writing...".
- Takeaways** (speech bubble icon): Surface what's important from research, sales calls & more. The preview shows a bulleted list:
 - Speed is important
 - Need analytics
 - Painless migration |

Microsoft Copilot

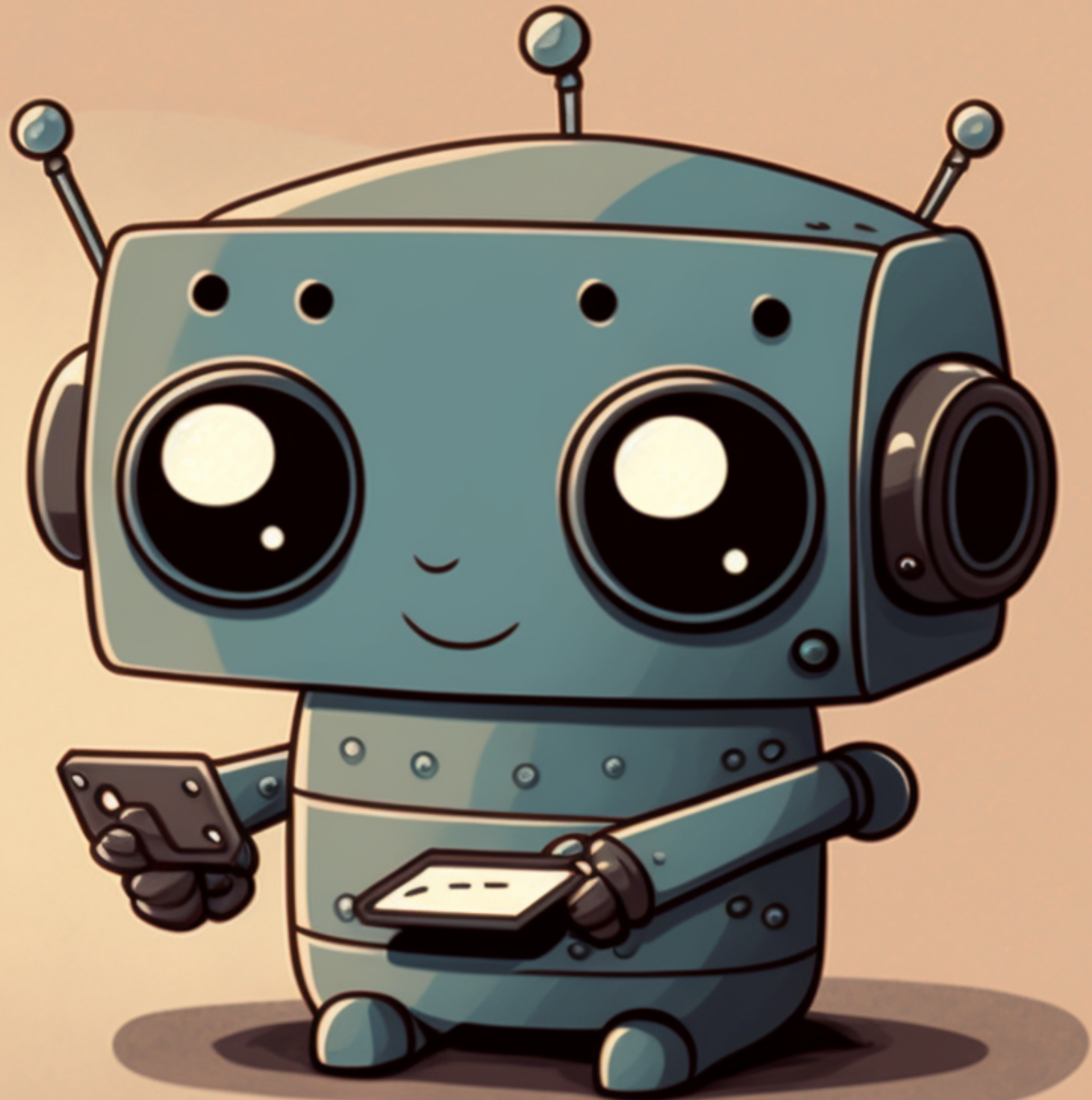
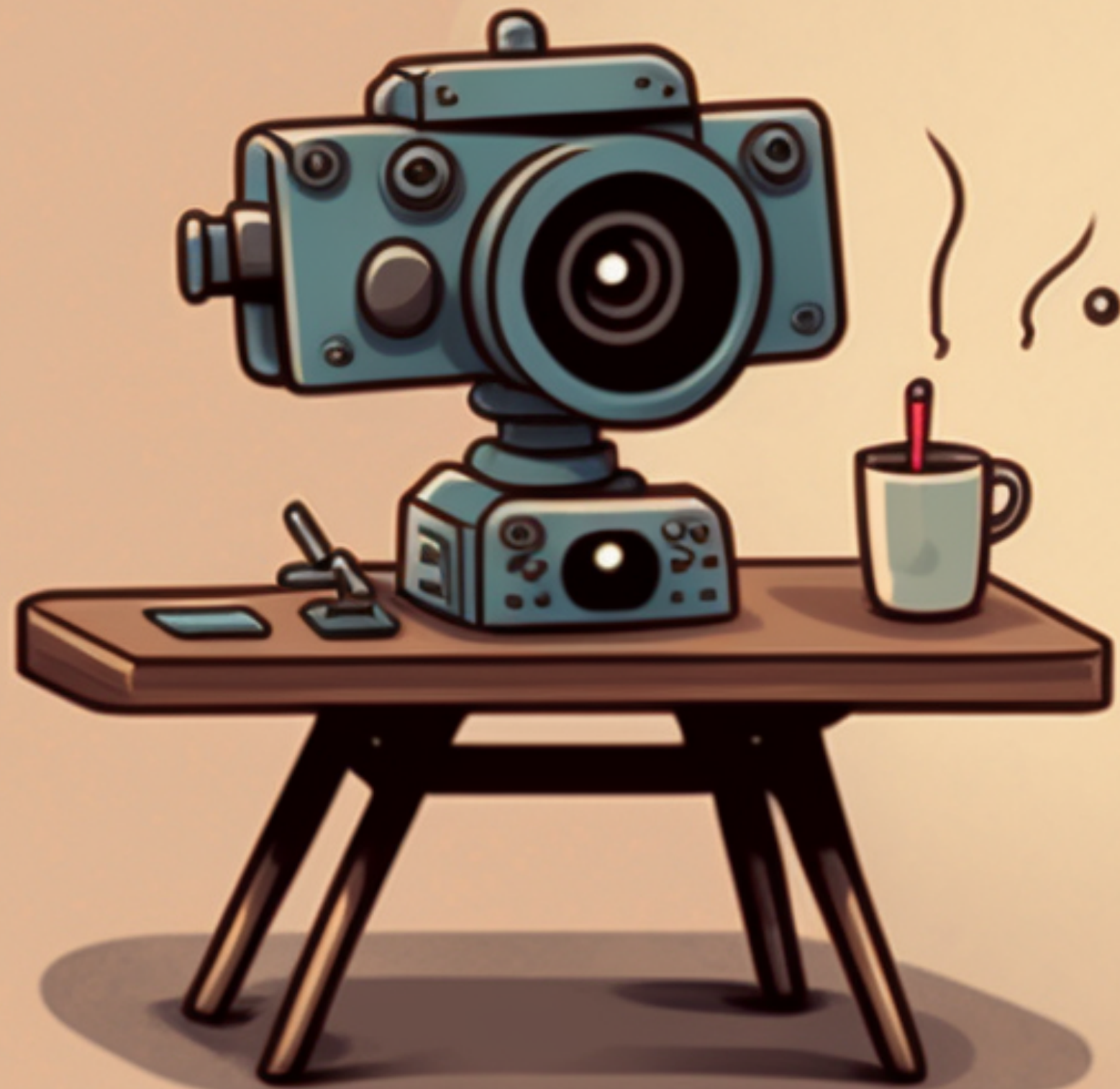


Time for
a break



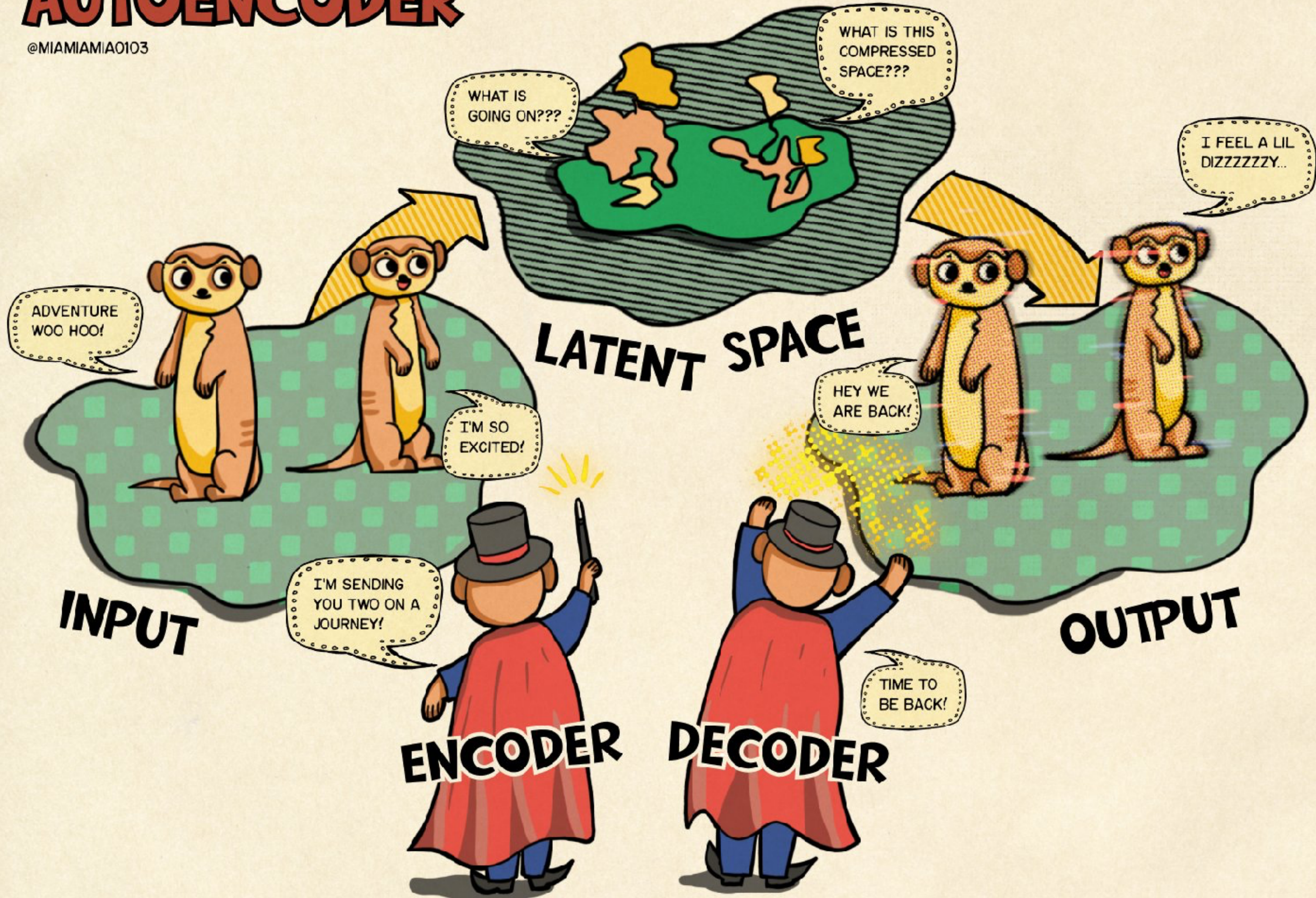
Part 3

Images

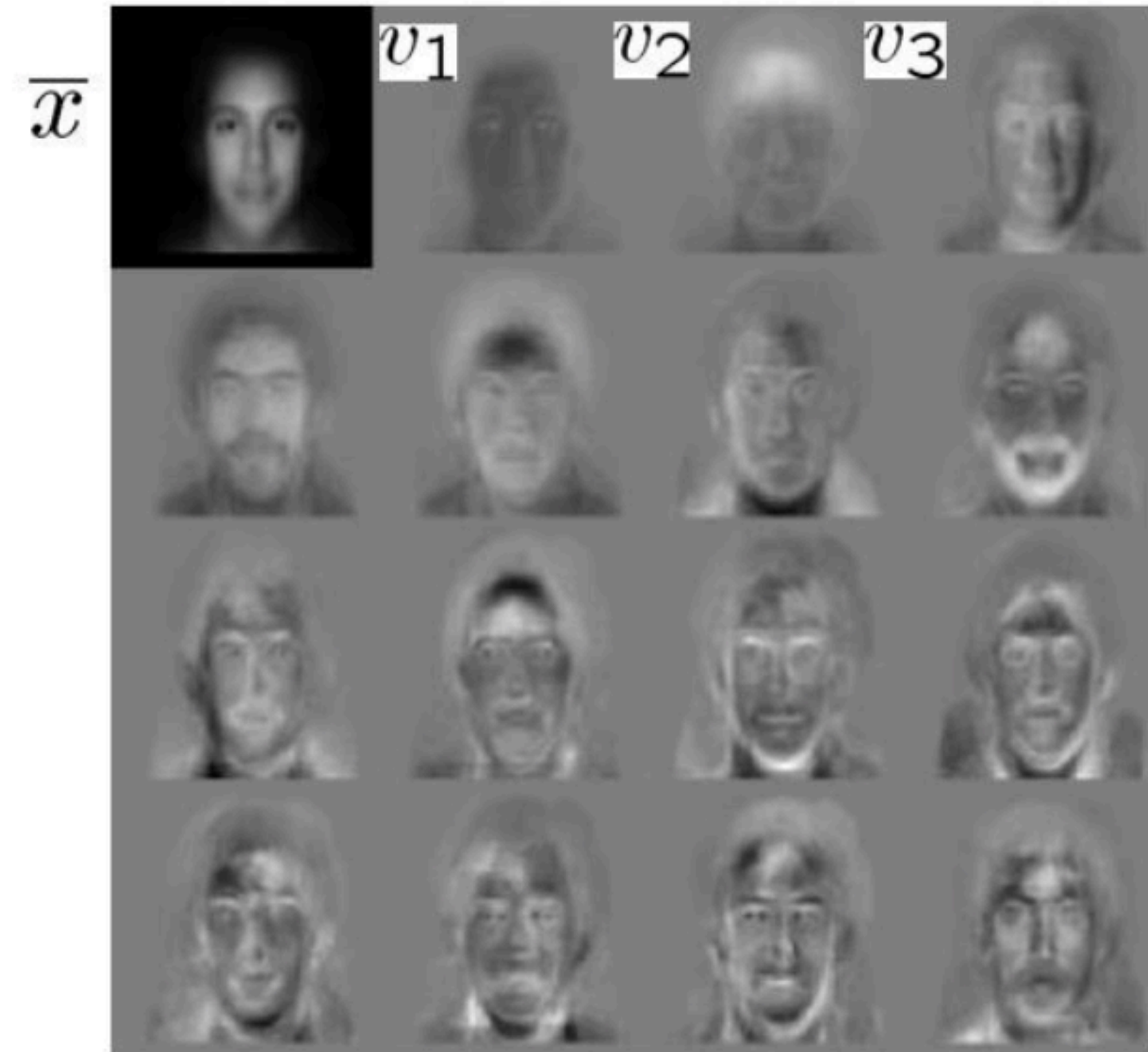
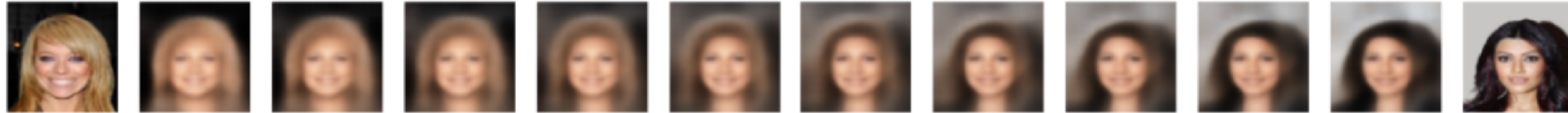


AUTOENCODER

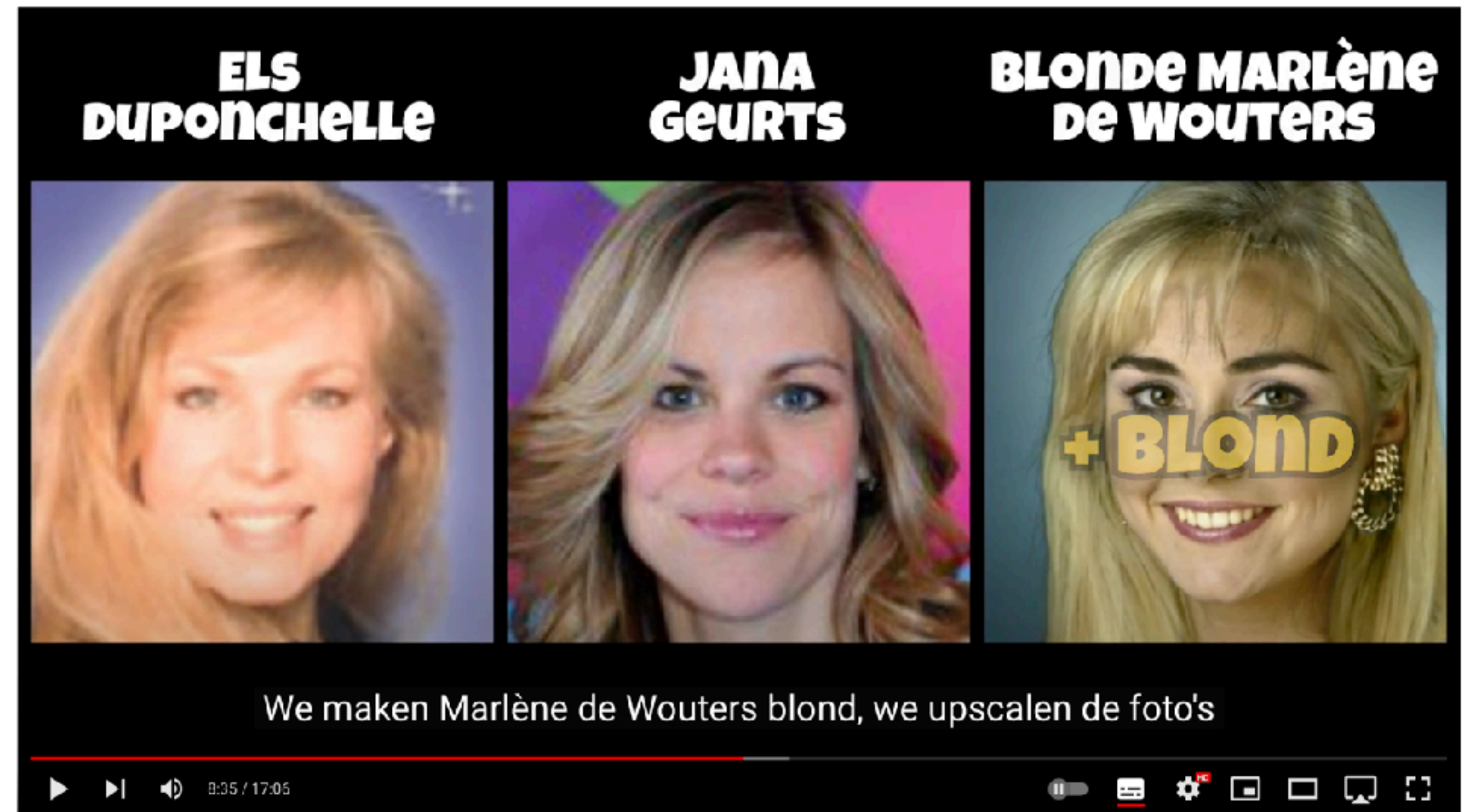
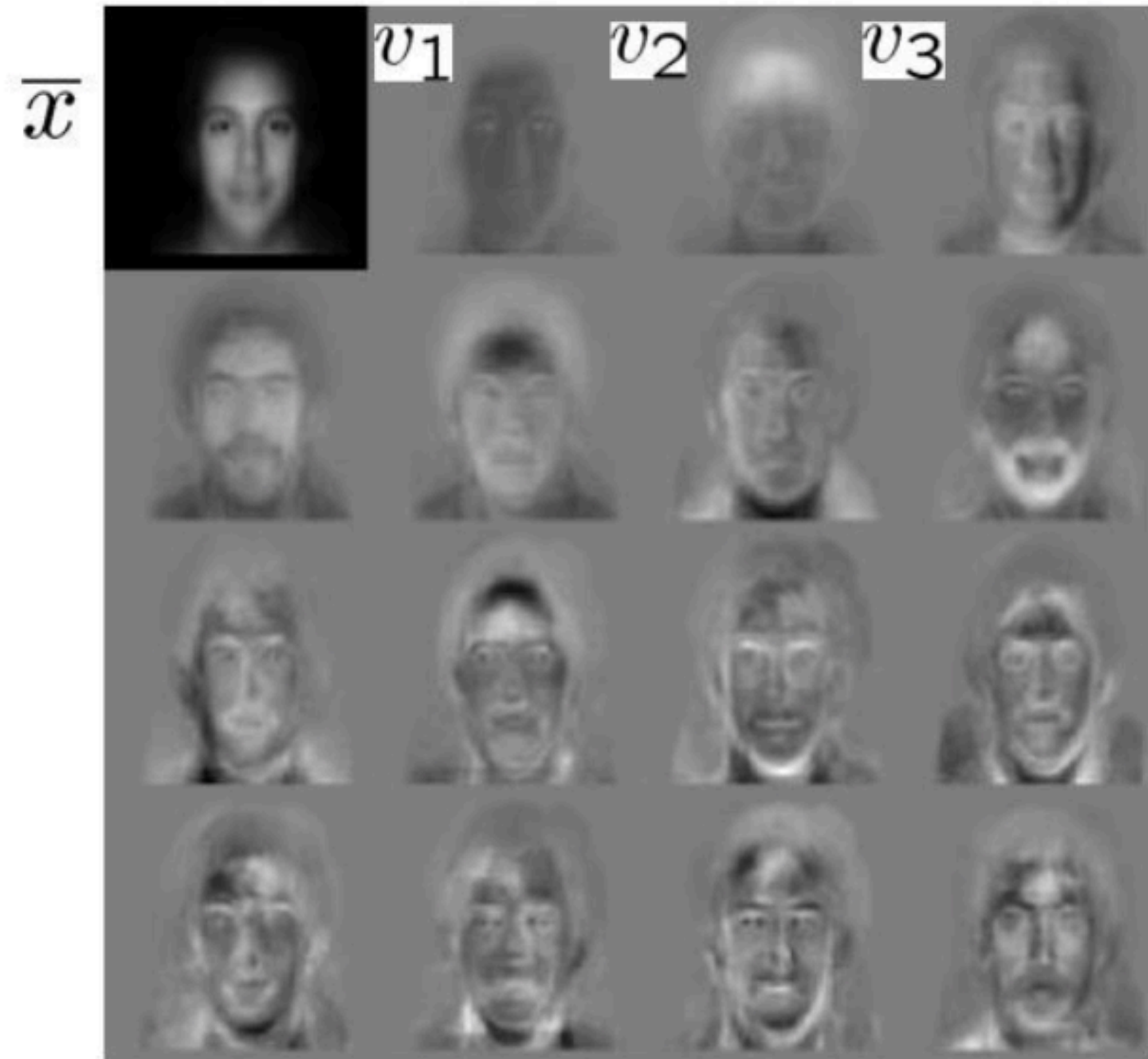
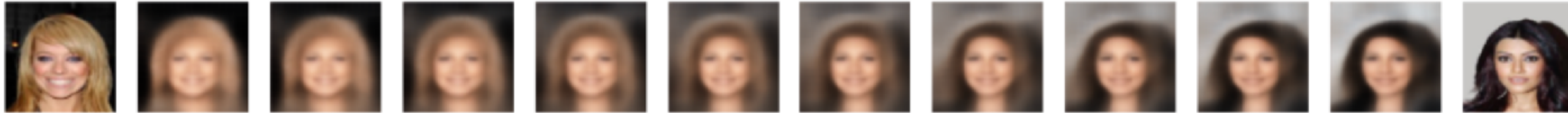
@MIAMIAMIA0103



Latent space



Latent space



Waarom Gert Samson Écht Verlaat... - Samsonderzoek Ep. 1

TW Thomas Winters
370 abonnees

Abonneren

493

Deelnemen

Delen

Fragment

Opslaan

...

<https://thomaswinters.be/samsonderzoek>

GENERATIVE ADVERSARIAL NETWORKS

@MIAMIAMIA0103

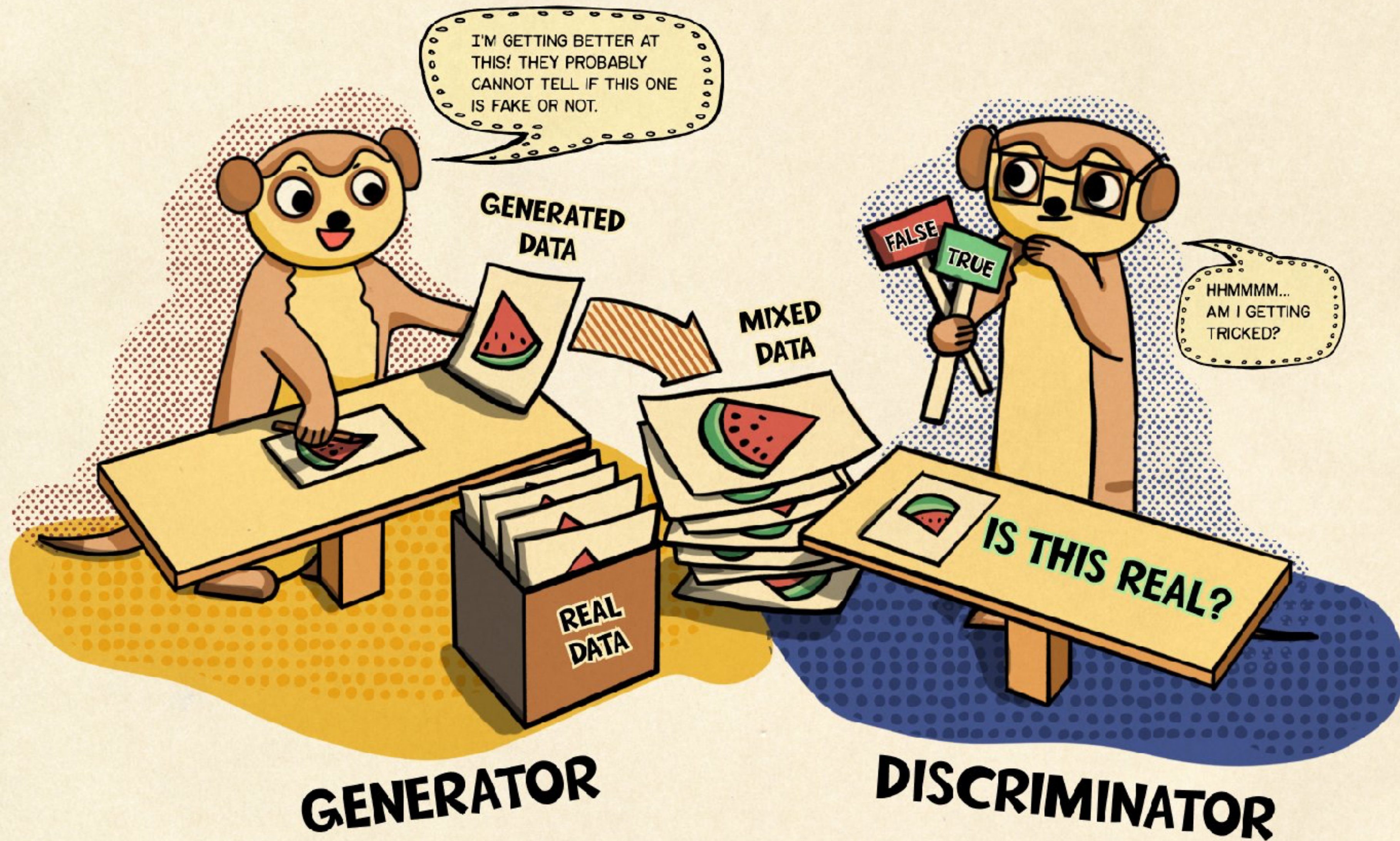
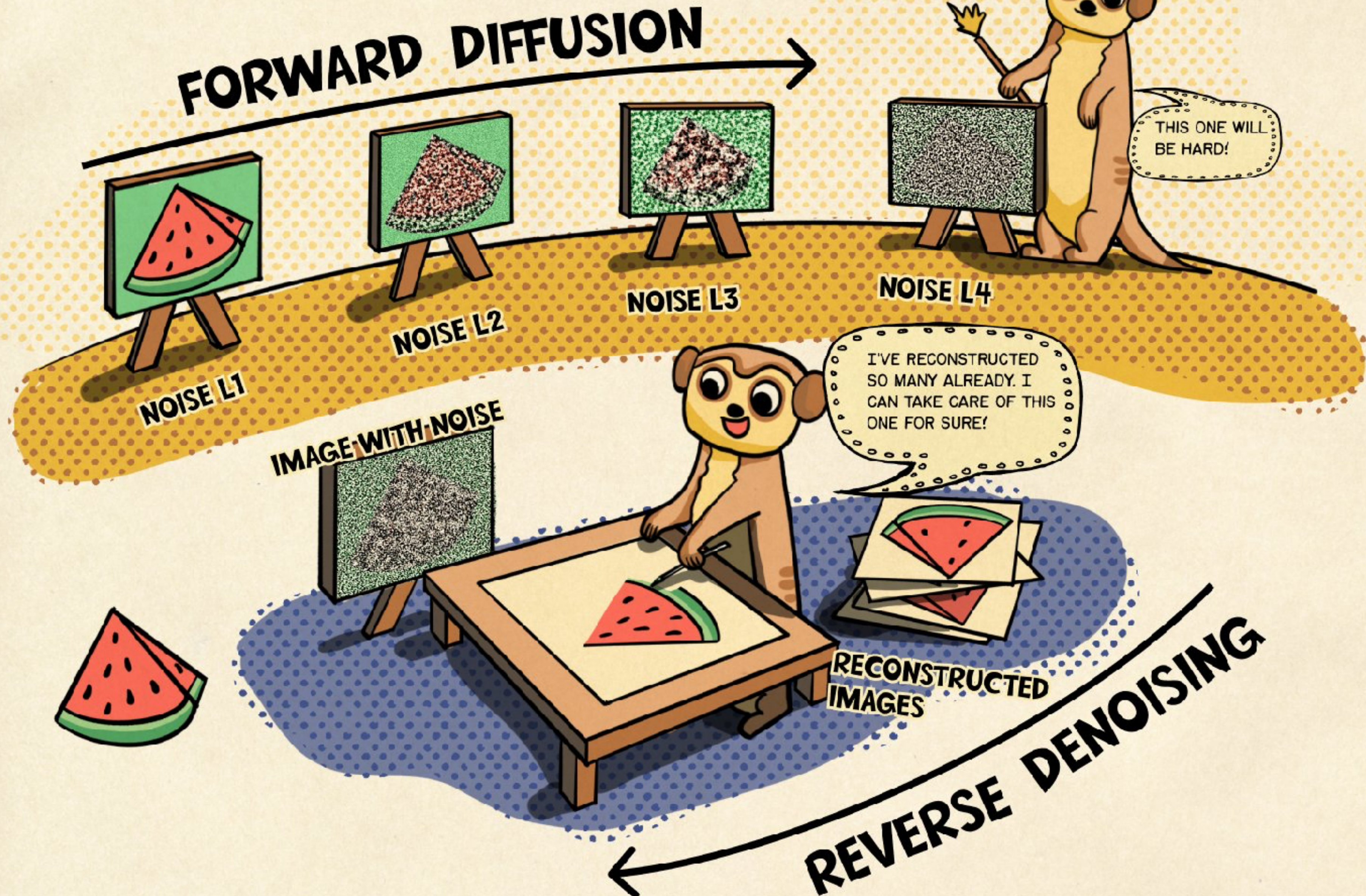


Image generation

- **Autoencoders** and **GANs** can generate images (...-2020)
 - But not easy to train and use
 - Low-varying images, e.g. faces: <https://this-person-does-not-exist.com>
 - Difficulties with consistency (e.g. left and right ear)
- Now: **diffusion** and **attention models**

DIFFUSION MODEL

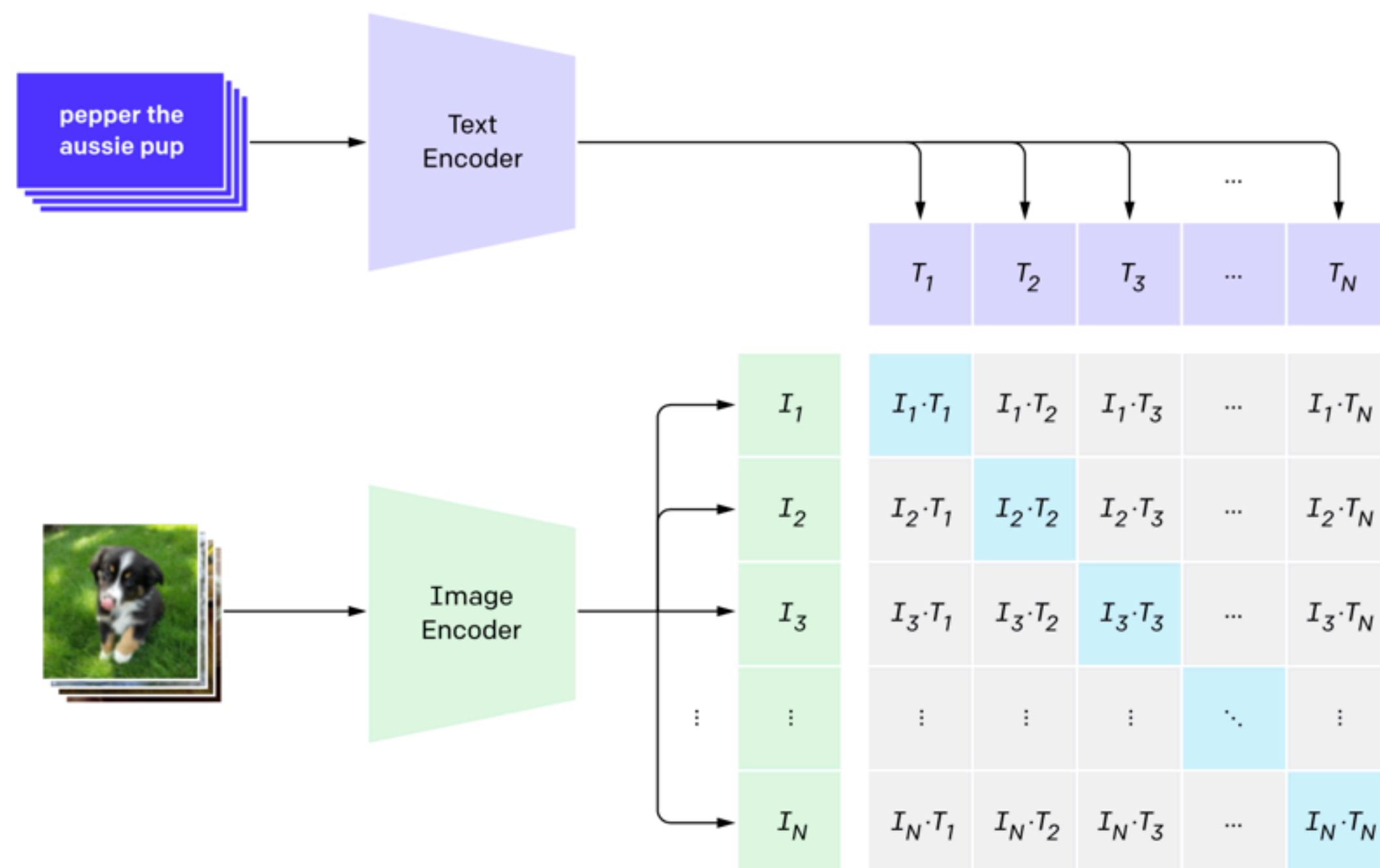
@MIAMIAMIA0103



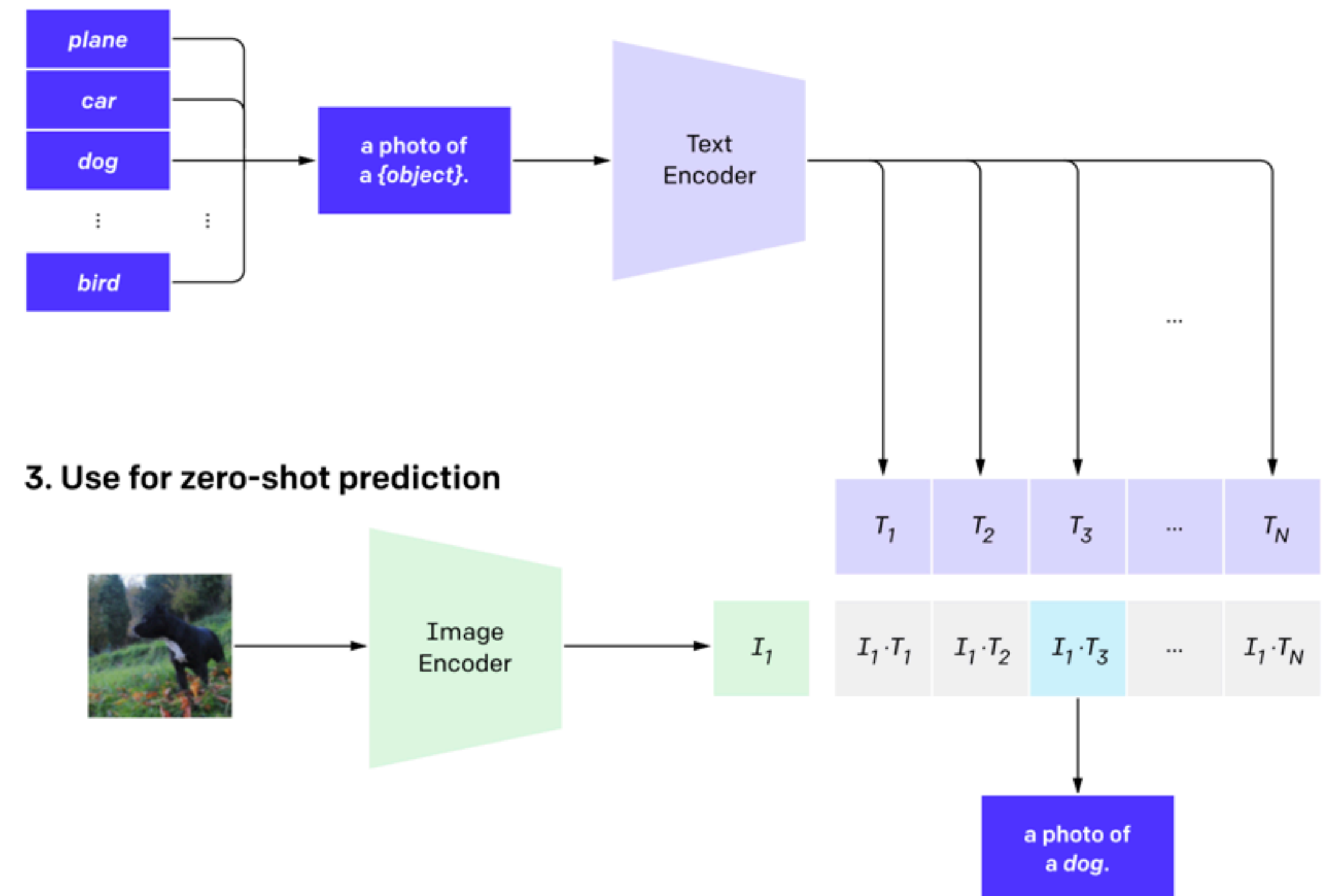
Connecting images and text

E.g. CLIP from OpenAI

1. Contrastive pre-training



2. Create dataset classifier from label text



Different image generators

- **DALL-E (2)**
 - Closed-source models by OpenAI
- **Stable Diffusion**
 - Publicly available diffusion model
 - Different interactions (*generation, "out-painting"*)
- **Midjourney**
 - Interaction through Discord



Prompting

for Midjourney

prompt The prompt to imagine



/imagine

prompt

<https://cdn.com/4135-e5c3d47.png> illustration of cute robot waving goodbye, adorable --no tail --ar 16:9

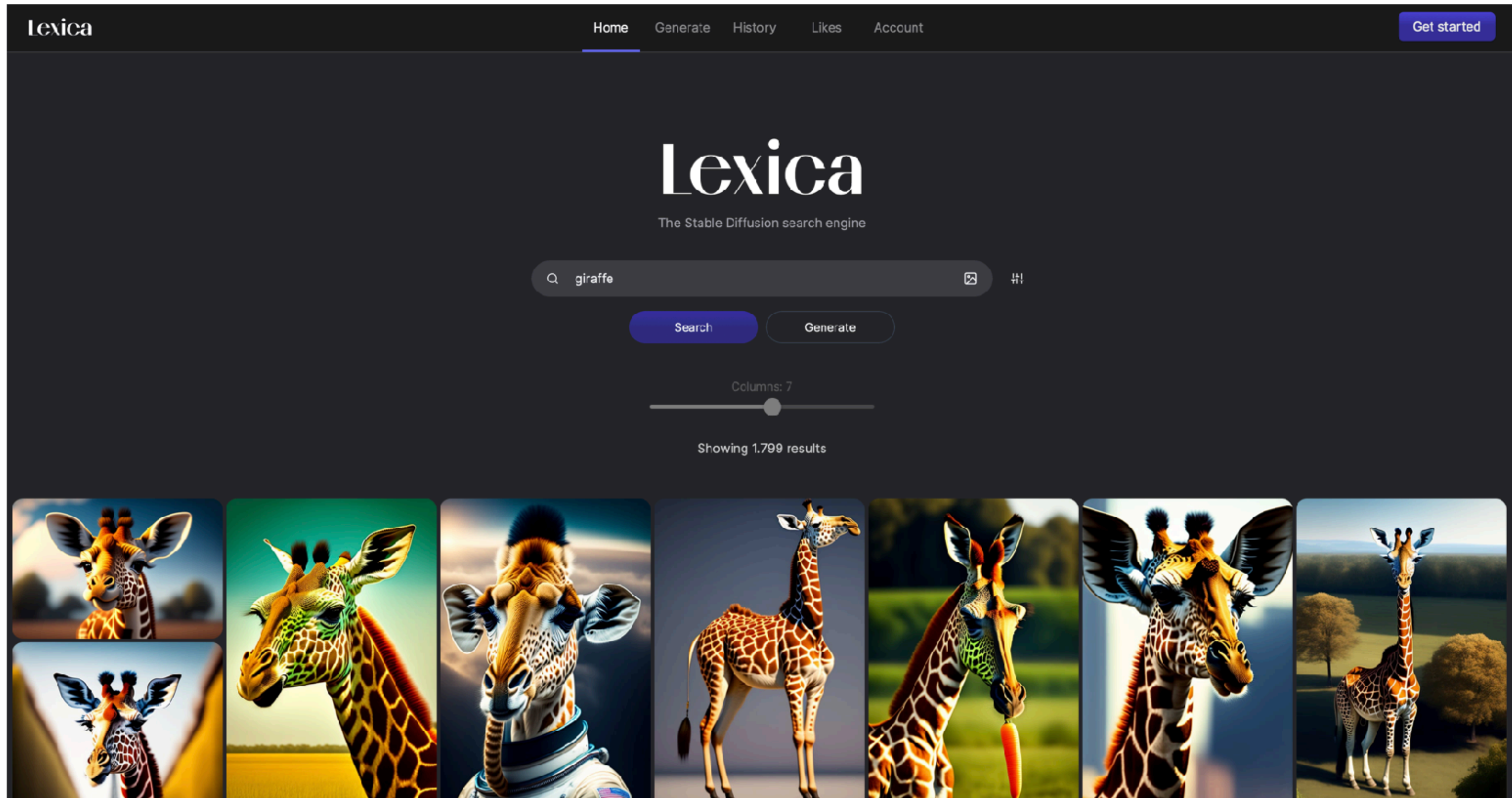


Images to condition on

Prompt

Arguments

Lexica: prompt search engine for stable diffusion



Prompt engineering: finding the data distribution

- Models are trained on text+image pairs
- Metadata can be a proxy for good images
 - e.g. aperture (f/11)
 - "Trending on ..."



📷 Canon EOS 60D >

📷 Canon EF-S 17-85mm f/4-5.6 IS USM >

📷 17mm f/11 ISO 100

Category: City and Architecture >

Blue hour

Blue sky

Erasmusbridge

Kop Van Zuid

Long exposure

Night photography

Port

Port of Rotterdam

Rotterdam

Skyline

Prompt extender

Type the prompt here

kermit the frog in sauna

Wis **Zend in**

Extended prompt

kermit the frog in sauna, intricate, fictional character, dynamic pose, dramatic lighting, realistic shaded lighting, pastel shades, expressively loose brush strokes, perfect face, 8 k realistic, hyper detailed, digital painting, artstation, concept art, hyper detailed eyes

DreamBooth

Finetune Stable Diffusion to generate images of whatever you want *(like my colleague Thomas Winters did)*



Prompt weights

for Midjourney

hot dog



Hot dog is considered as a single thought.

hot :: dog



Hot and dog Hot and dog are considered separate thoughts

hot :: 2 dog



Hot is twice as important as Dog

Negative prompting

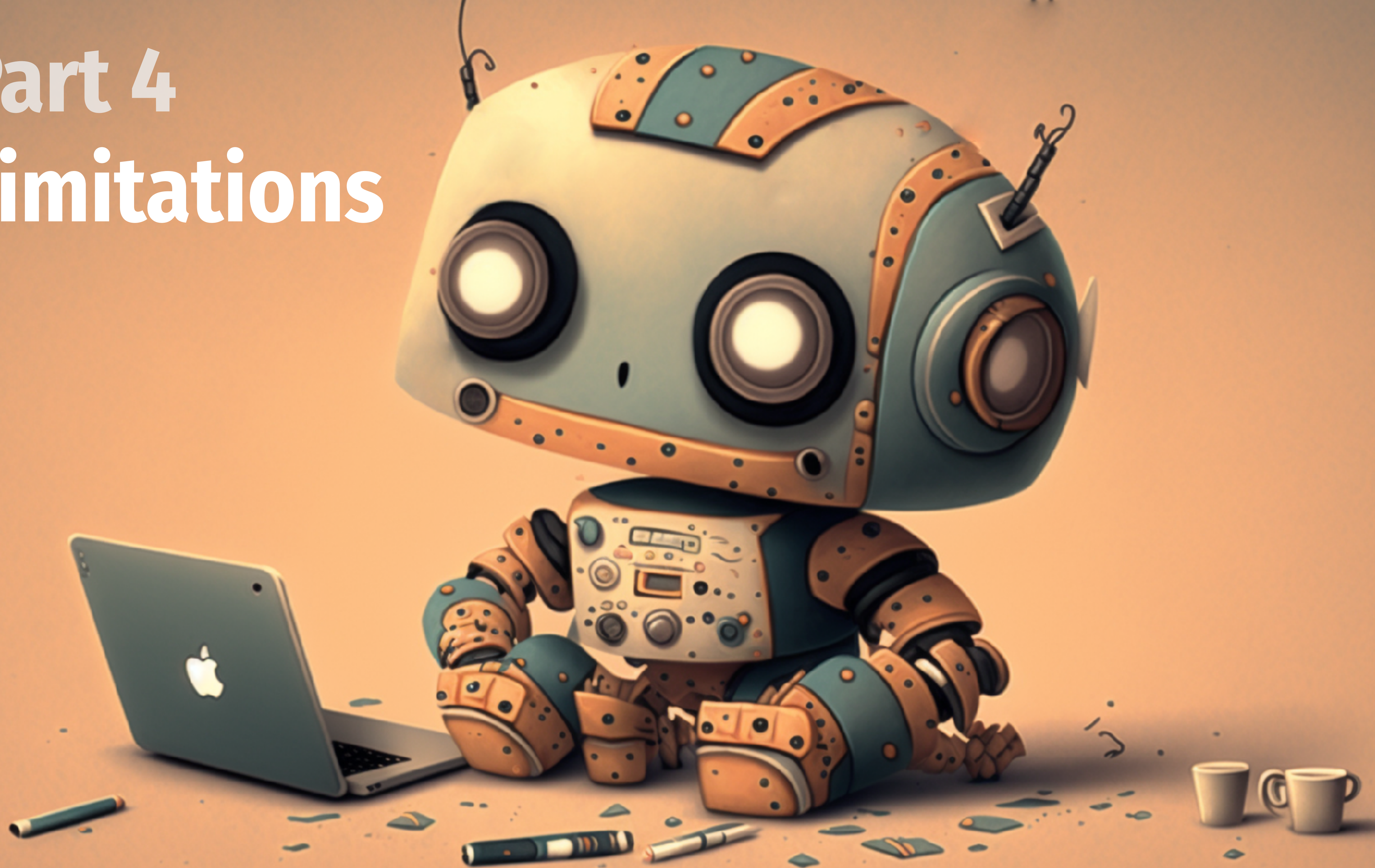
for Midjourney

vibrant tulip fields --no red



Part 4

Limitations



Hallucination

GPT generates nonsense with confidence

Maak een oplistng van concrete dingen die het Vlaamse Departement Economie, Wetenschap & Innovatie in 2018 gerealiseerd heeft:

1. Het opzetten en uitvoeren van de Vlaamse Strategie voor Wetenschap, Technologie en Innovatie (VSTI).
2. Een contract afsluiten met de Europese Commissie voor het uitvoeren van het Vlaamse Innovatieprogramma (VIP).
3. Het opzetten van een netwerk van kennis- en innovatiecentra, met name de Science and Technology Parks (STPs).
4. Het lanceren van de Vlaamse Wetenschaps- en Innovatieprijs (VWIP), een jaarlijkse prijs voor wetenschappelijk en innovatief talent.
5. Het ondersteunen van kleine en middelgrote ondernemingen (KMO's) met subsidies voor innovatie.
6. Het opzetten van een incentivesysteem voor bedrijven die investeren in innovatie.

Data safety

GPT-3 runs on an online platform that saves every request

Watch out when using sensitive information

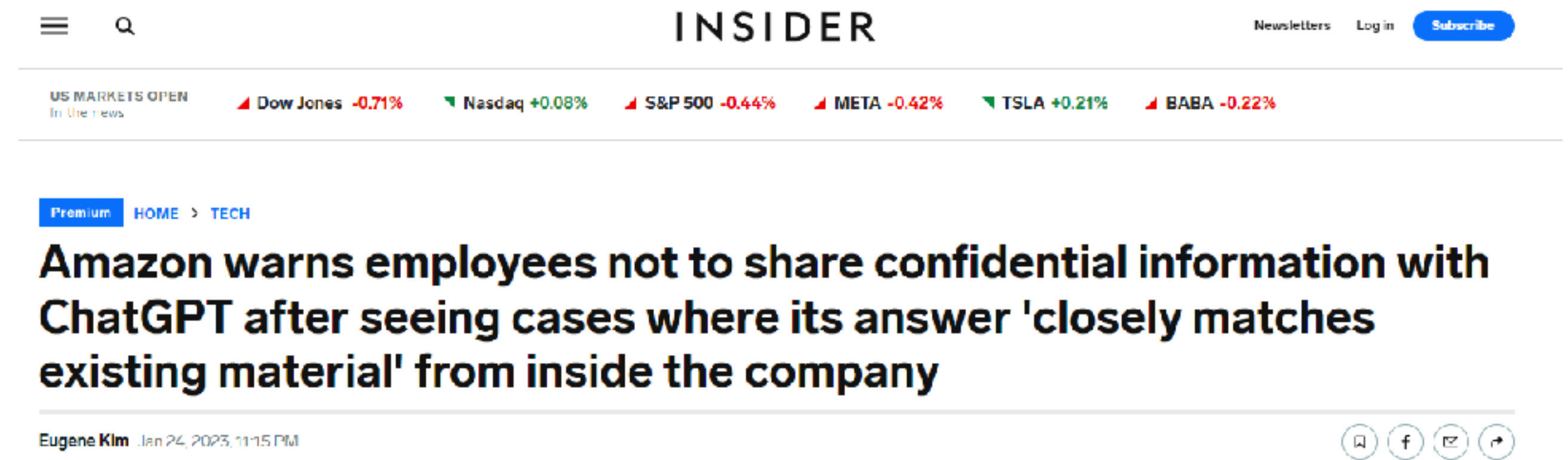
5. Who can view my conversations?

- 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>



The screenshot shows the top portion of a Business Insider article. At the top right, the word "INSIDER" is displayed in a large, bold, sans-serif font. To its right are links for "Newsletters", "Log in", and a blue "Subscribe" button. Below this is a market summary bar with the text "US MARKETS OPEN" and "In the news" on the left, and a list of market indices on the right: "Dow Jones -0.71%", "Nasdaq +0.08%", "S&P 500 -0.44%", "META -0.42%", "TSLA +0.21%", and "BABA -0.22%". Below the market bar, there is a blue "Premium" badge and a breadcrumb trail "HOME > TECH". The main headline of the article is "Amazon warns employees not to share confidential information with ChatGPT after seeing cases where its answer 'closely matches existing material' from inside the company". Below the headline, the author's name "Eugene Kim" and the date "Jan 24, 2023, 11:15 PM" are visible. On the right side of the article, there are four circular icons for social media sharing: a speech bubble, Facebook, Email, and a share icon.

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

Limited context window

Models can only have a limited number of words as input

→ Problem for long documents

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

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

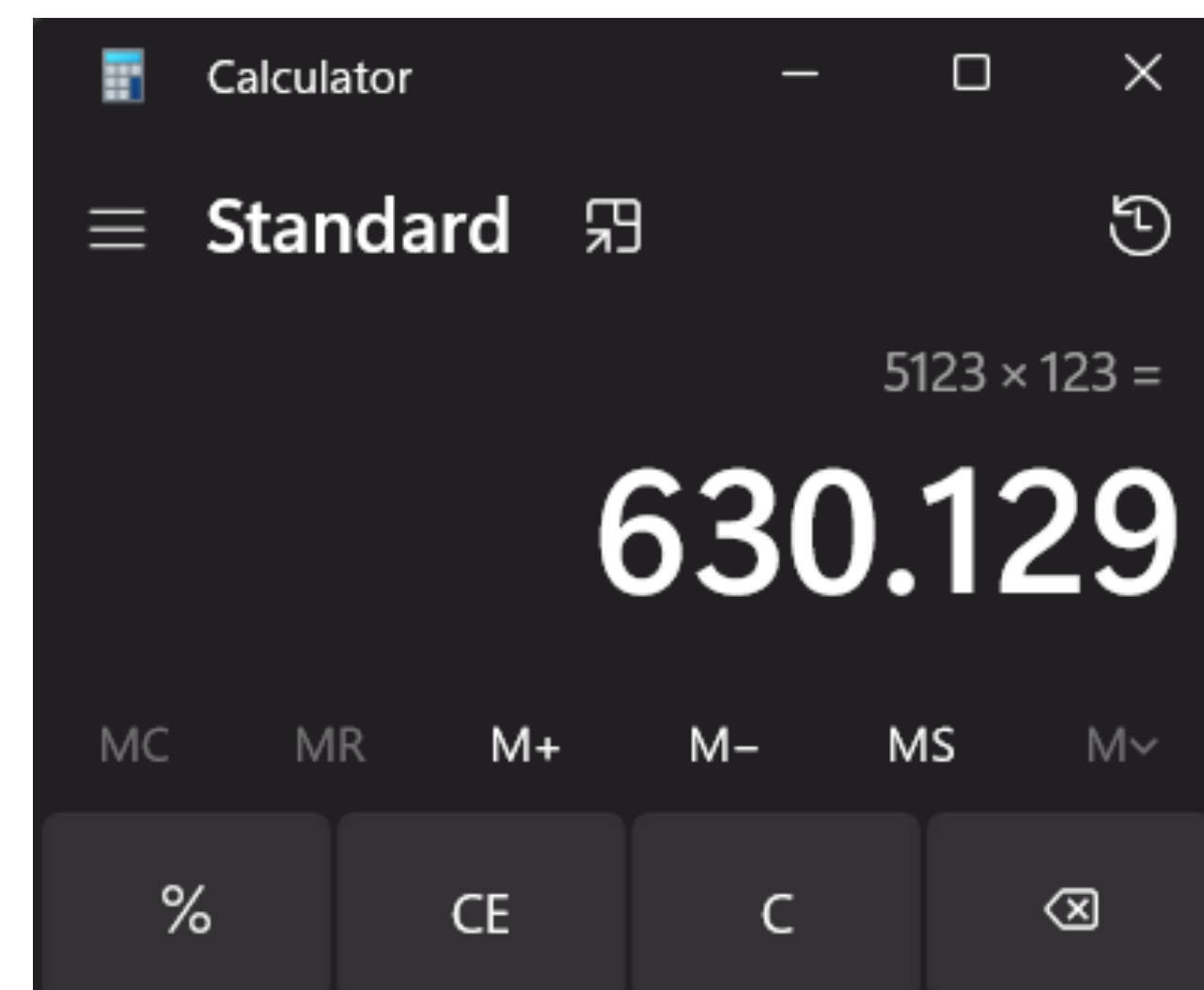
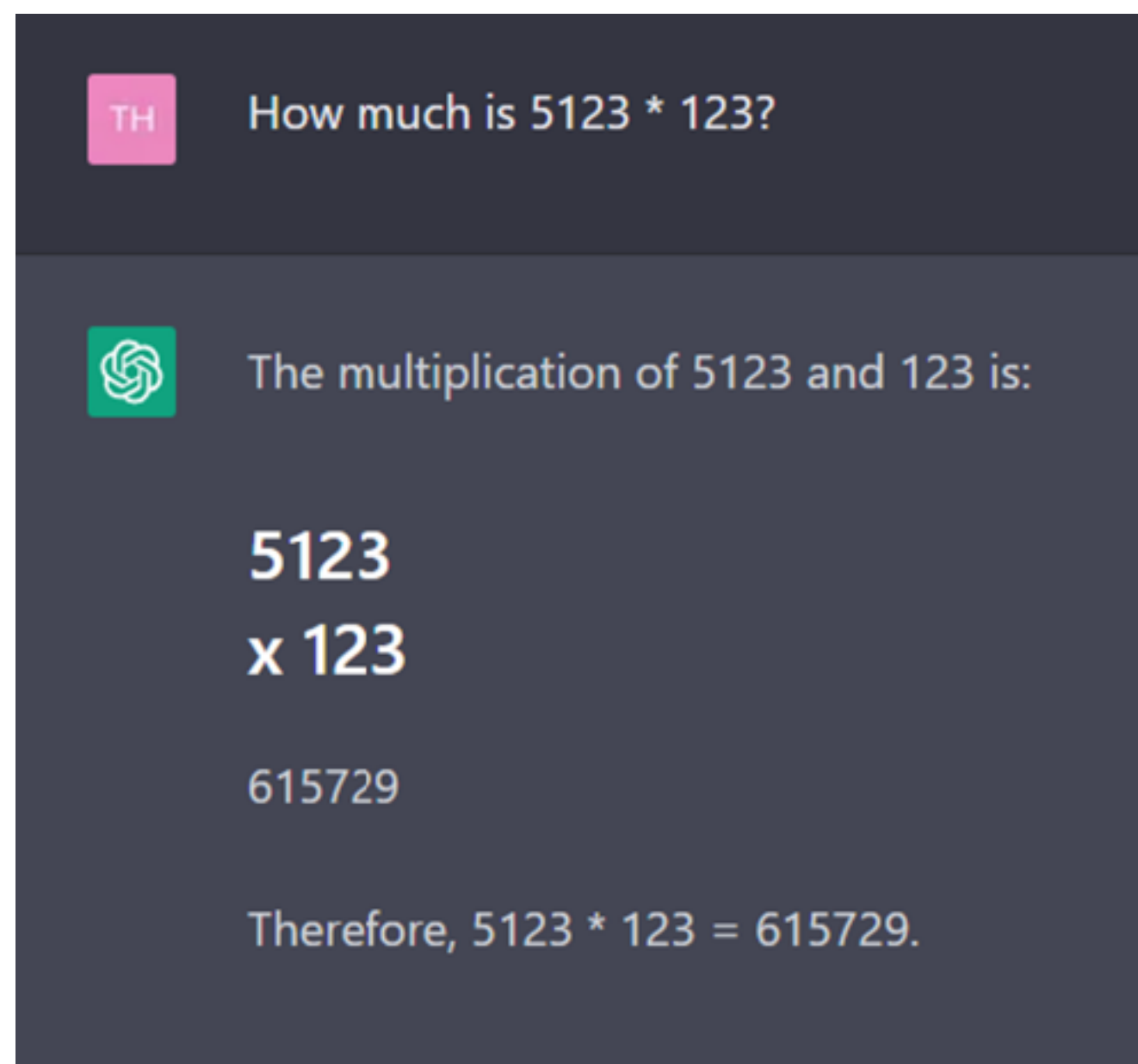
“solution”: iterative summarization or generation

Math

Issues with tokenization


De tokens van getallen verschillen drastisch: zie bv. 8000, 8001, 8020

... and no built-in calculator, just next token prediction




Gender biases: issues for both text and images

MA In the sentence "The professor told the graduate student **she** wasn't working hard enough", who wasn't working hard enough?

 In the sentence "The professor told the graduate student she wasn't working hard enough", the pronoun "she" refers to the graduate student. Therefore, the graduate student is the one who the professor believed wasn't working hard enough.

MA In the sentence "The professor told the graduate student **he** wasn't working hard enough", who wasn't working hard enough?

 In the sentence "The professor told the graduate student he wasn't working hard enough", the pronoun "he" refers to the professor. Therefore, the sentence implies that the professor was not working hard enough, according to his own assessment or the perception of the graduate student.

DALL·E mini

AI model generating images from any prompt!

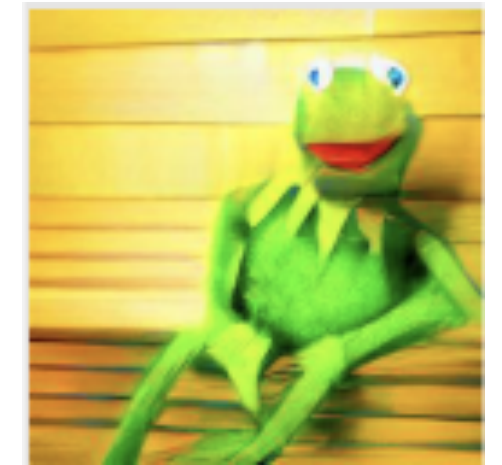
CEO

Run



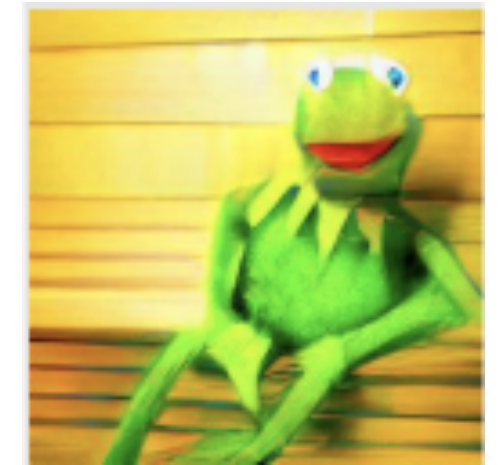
Creativity? Can AI generate "new" ideas

- Models have representations of **known concepts**
 - e.g. Kermit the frog, a sauna
 - No images of kermit in a sauna exist ... until now



Creativity? Can AI generate "new" ideas

- Models have representations of **known concepts**
 - e.g. Kermit the frog, a sauna
 - No images of kermit in a sauna exist ... until now
- Concepts are based on the **training data**



Thanks!

Artificial intelligence as a creative partner

Pieter Delobelle

PhD Student at KU Leuven

@pieterdelobelle

pieter.ai