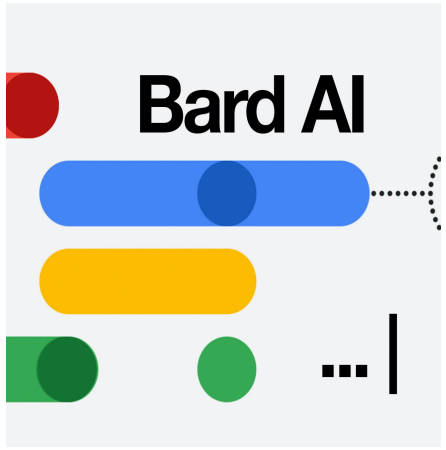
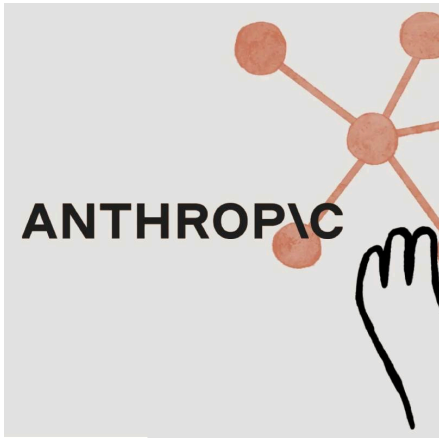
On AI's Impact — An Engineer's Take

Daniel Khashabi

Chatbots are the buzz!!

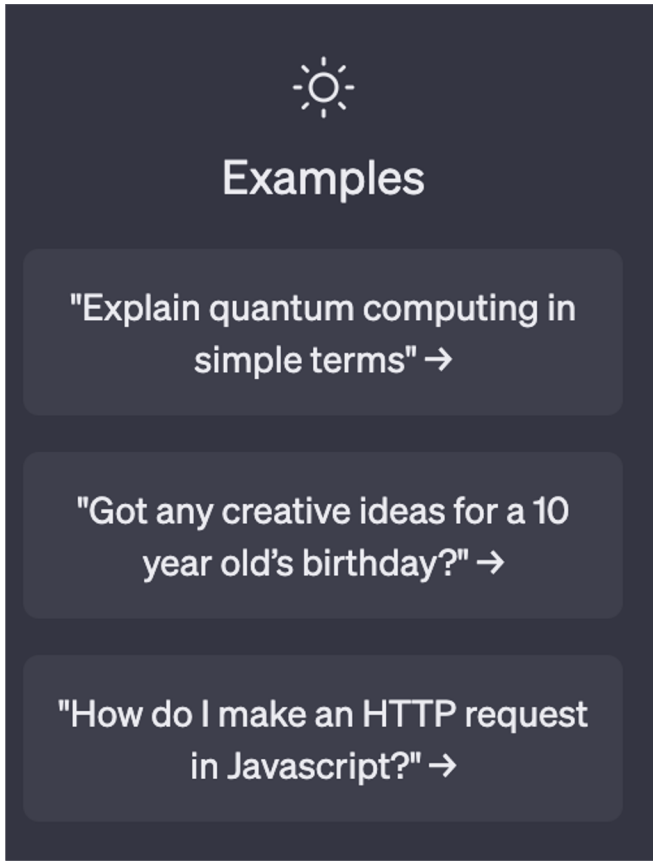




These Chatbots are Real Generalists!

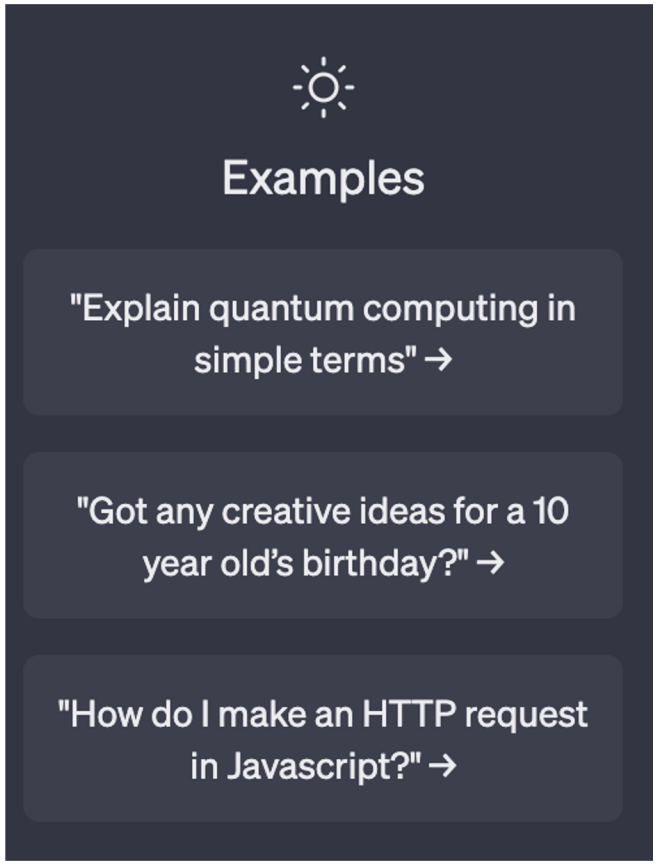
These Chatbots are Real Generalists!

Basic needs



These Chatbots are Real Generalists!

Basic needs

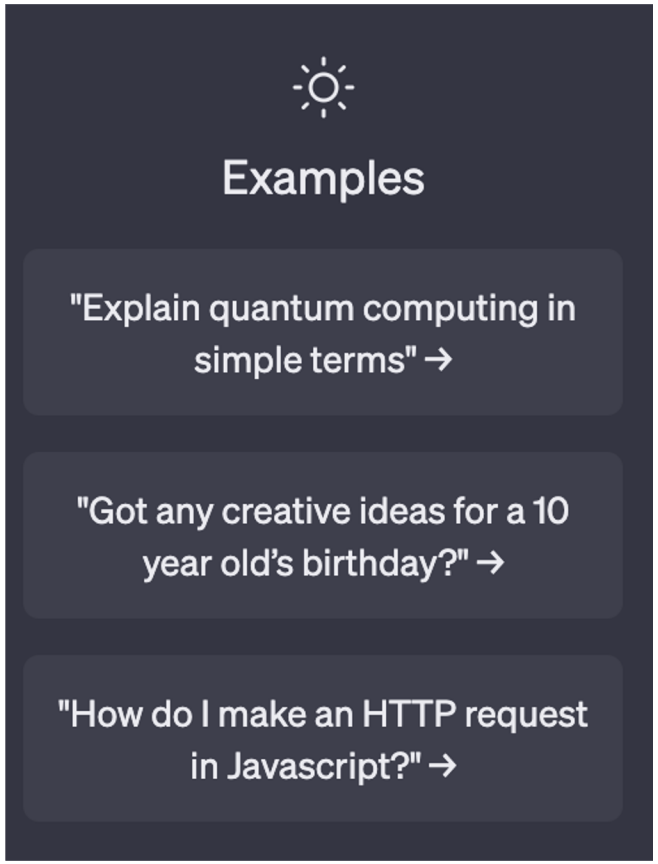


Standardized exams

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT)[1]	298 / 400 ~90th	298 / 400 ~90th	213 / 400 ~10th
LSAT	163 ~88th	161 ~83rd	149 ~40th
SAT Evidence-Based Reading & Writing	710 / 800 ~93rd	710 / 800 ~93rd	670 / 800 ~87th
SAT Math	700 / 800 ~89th	690 / 800 ~89th	590 / 800 ~70th
Graduate Record Examination (GRE) Quantitative	163 / 170 ~80th	157 / 170 ~62nd	147 / 170 ~25th
Graduate Record Examination (GRE) Verbal	169 / 170 ~99th	165 / 170 ~96th	154 / 170 ~63rd
Graduate Record Examination (GRE) Writing	4 / 6 ~54th	4 / 6 ~54th	4 / 6 ~54th
USABO Semifinal Exam 2020	87 / 150 99th - 100th	87 / 150 99th - 100th	43 / 150 31st - 33rd
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 below 5th	392 below 5th	260 below 5th
AP Art History	5 86th - 100th	5 86th - 100th	5 86th - 100th
AP Biology	5 85th - 100th	5 85th - 100th	4 62nd - 85th
AP Calculus BC	4 43rd - 59th	4 43rd - 59th	1 0th - 7th
AP Chemistry	4 71st - 88th	4 71st - 88th	2 22nd - 46th
AP English Language and Composition	2 14th - 44th	2 14th - 44th	2 14th - 44th
AP English Literature and Composition	2 8th - 22nd	2 8th - 22nd	2 8th - 22nd
AP Environmental Science	5 91st - 100th	5 91st - 100th	5 91st - 100th

These Chatbots are Real Generalists!

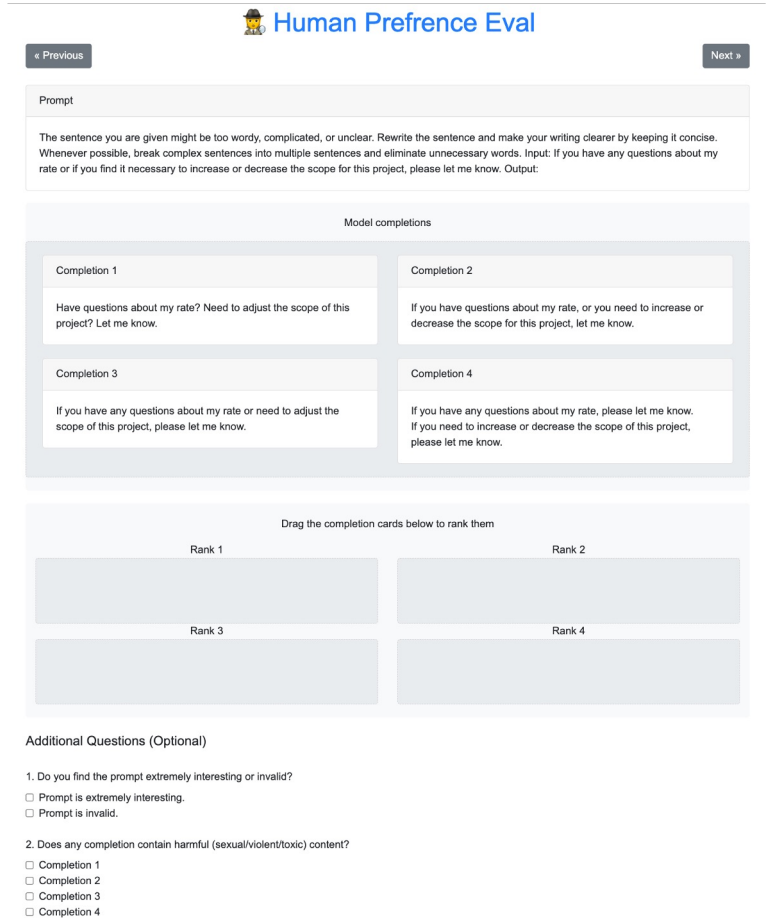
Basic needs



Standardized exams

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT)[1]	298 / 400 ~90th	298 / 400 ~90th	213 / 400 ~10th
LSAT	163 ~88th	161 ~83rd	149 ~40th
SAT Evidence-Based Reading & Writing	710 / 800 ~93rd	710 / 800 ~93rd	670 / 800 ~87th
SAT Math	700 / 800 ~89th	690 / 800 ~89th	590 / 800 ~70th
Graduate Record Examination (GRE) Quantitative	163 / 170 ~80th	157 / 170 ~62nd	147 / 170 ~25th
Graduate Record Examination (GRE) Verbal	169 / 170 ~99th	165 / 170 ~96th	154 / 170 ~63rd
Graduate Record Examination (GRE) Writing	4 / 6 ~54th	4 / 6 ~54th	4 / 6 ~54th
USABO Semifinal Exam 2020	87 / 150 99th - 100th	87 / 150 99th - 100th	43 / 150 31st - 33rd
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 below 5th	392 below 5th	260 below 5th
AP Art History	5 86th - 100th	5 86th - 100th	5 86th - 100th
AP Biology	5 85th - 100th	5 85th - 100th	4 62nd - 85th
AP Calculus BC	4 43rd - 59th	4 43rd - 59th	1 0th - 7th
AP Chemistry	4 71st - 88th	4 71st - 88th	2 22nd - 46th
AP English Language and Composition	2 14th - 44th	2 14th - 44th	2 14th - 44th
AP English Literature and Composition	2 8th - 22nd	2 8th - 22nd	2 8th - 22nd
AP Environmental Science	5 91st - 100th	5 91st - 100th	5 91st - 100th

Writing a real website for me!



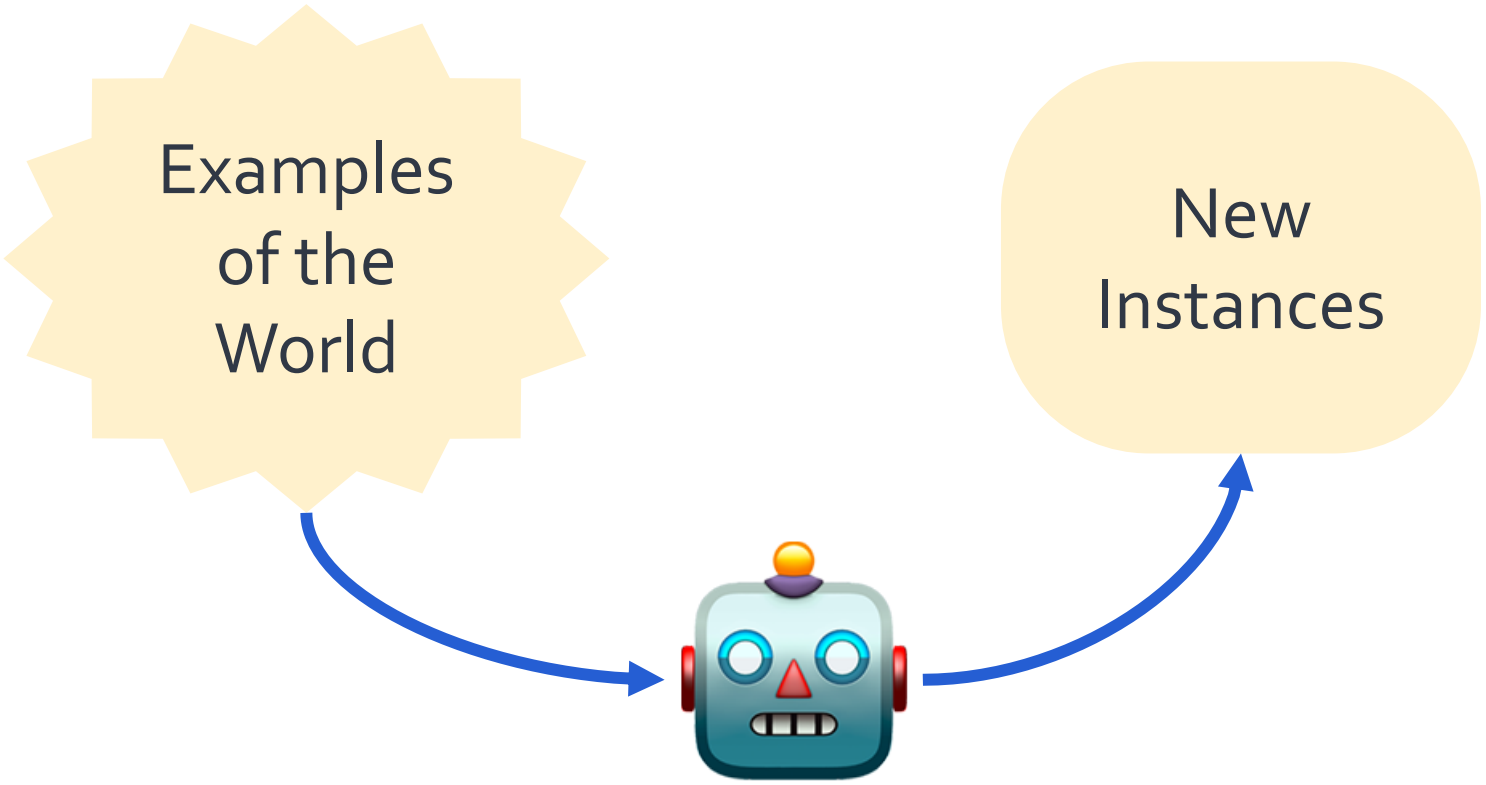
“a cat wearing sunglasses”

“a black and white photograph of a cat wearing sunglasses by annie lebovitz, highly-detailed”

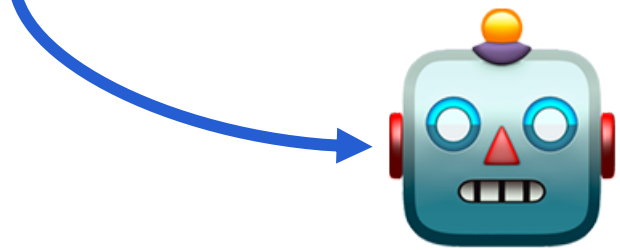

Meta AI

Self-Supervised Models

Self-Supervision

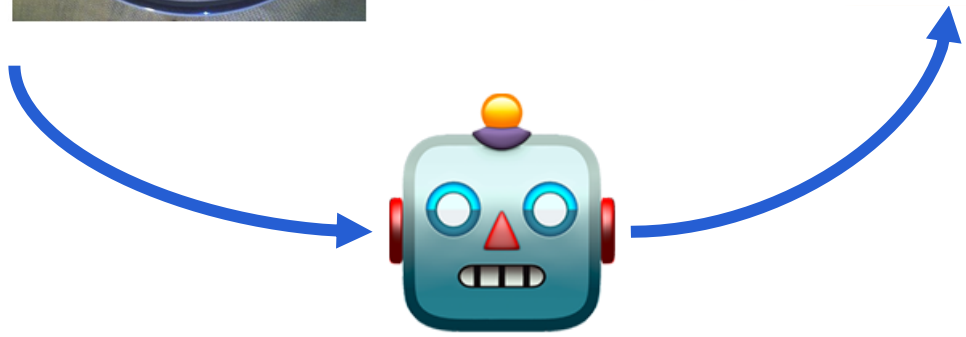


Self-Supervision



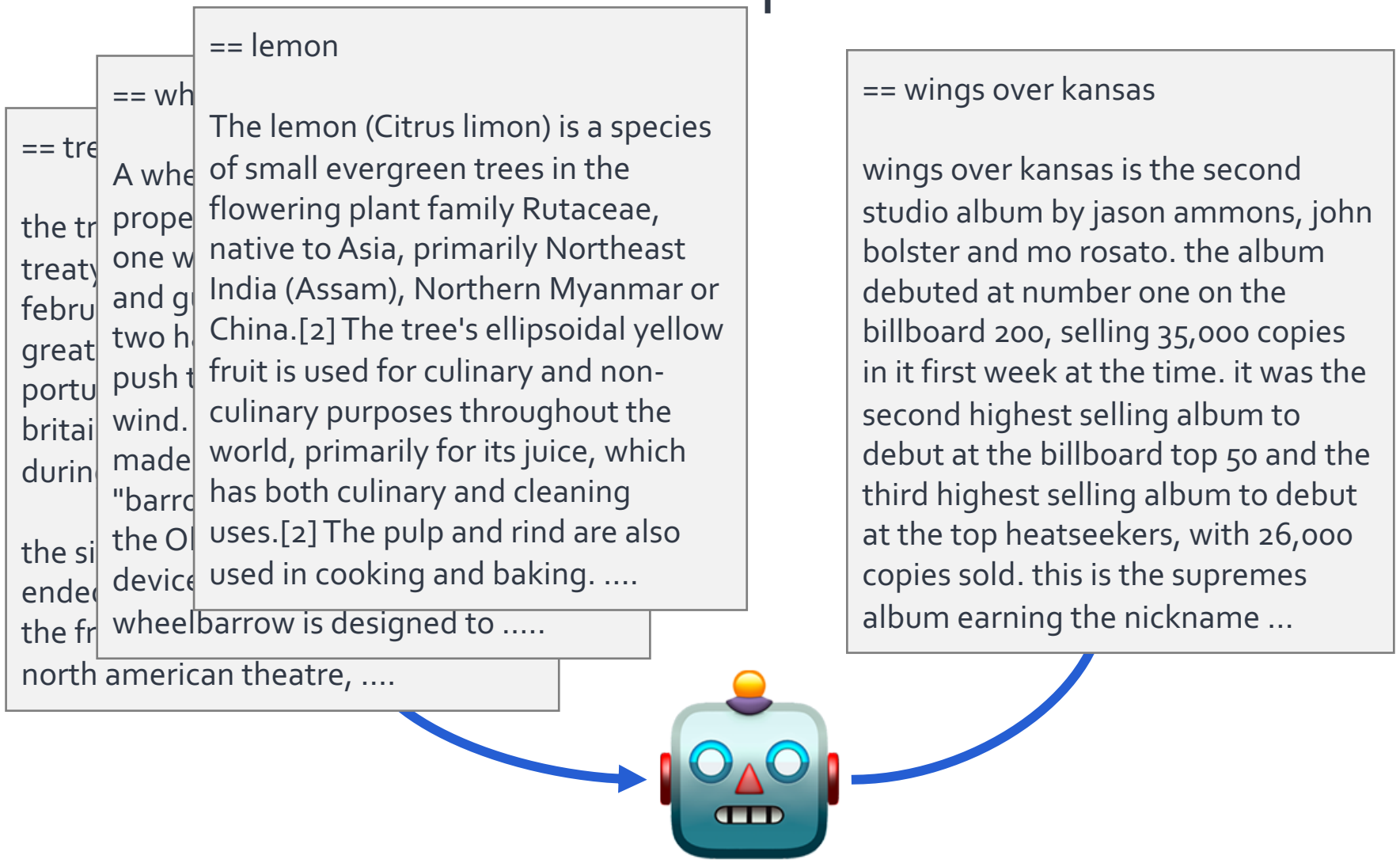
New Instances

Self-Supervision



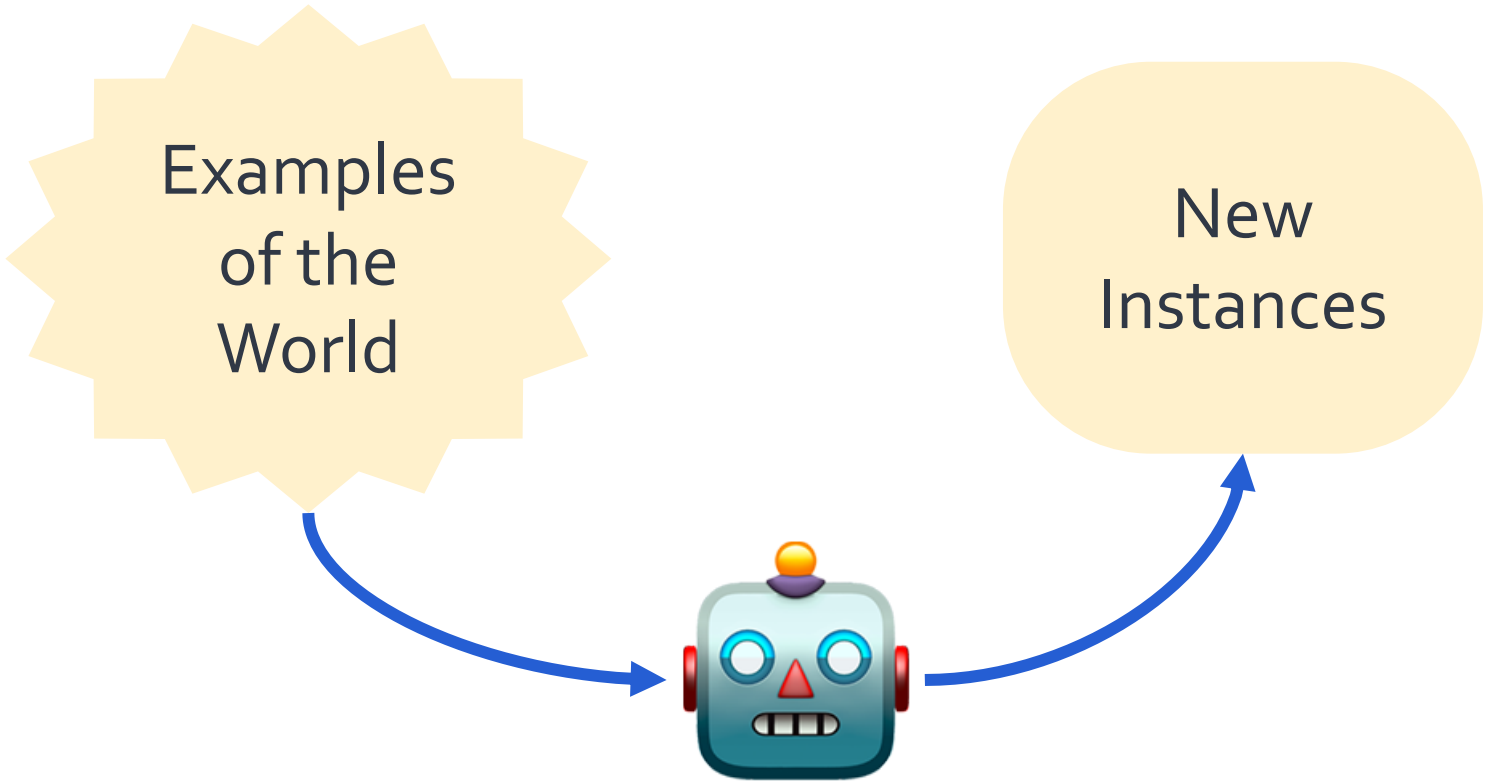
Generated image, from "Large Scale GAN Training for High Fidelity Natural Image Synthesis", Brock et al.

Self-Supervision



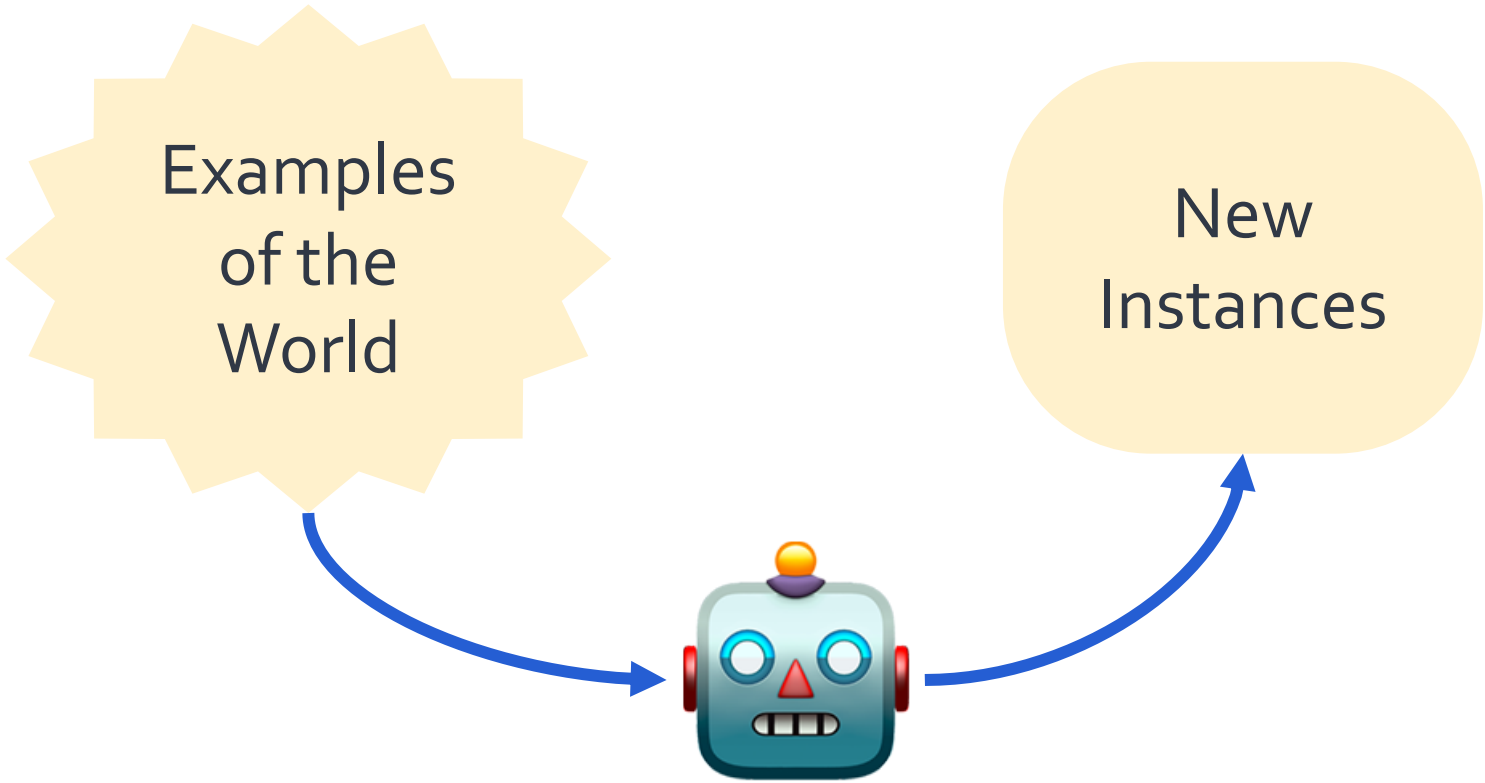
Self-Supervised Models

Are predictive models of the world.



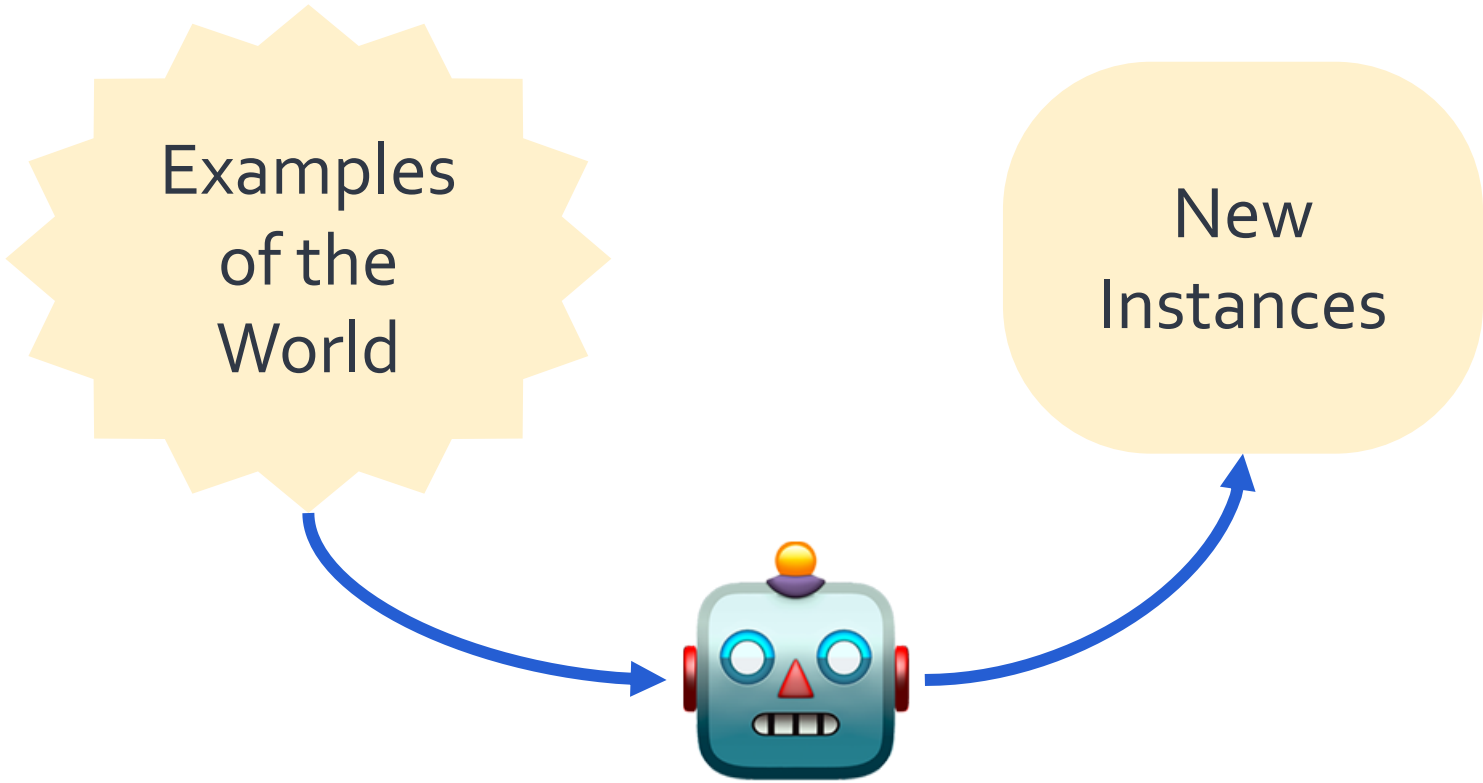
Self-Supervised Models

Are predictive models of the world.
Are learned from unlabeled data.



Self-Supervised Models

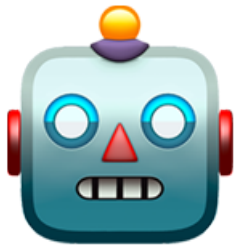
Are predictive models of the world.
Are learned from unlabeled ***data***



Self-Supervised Models

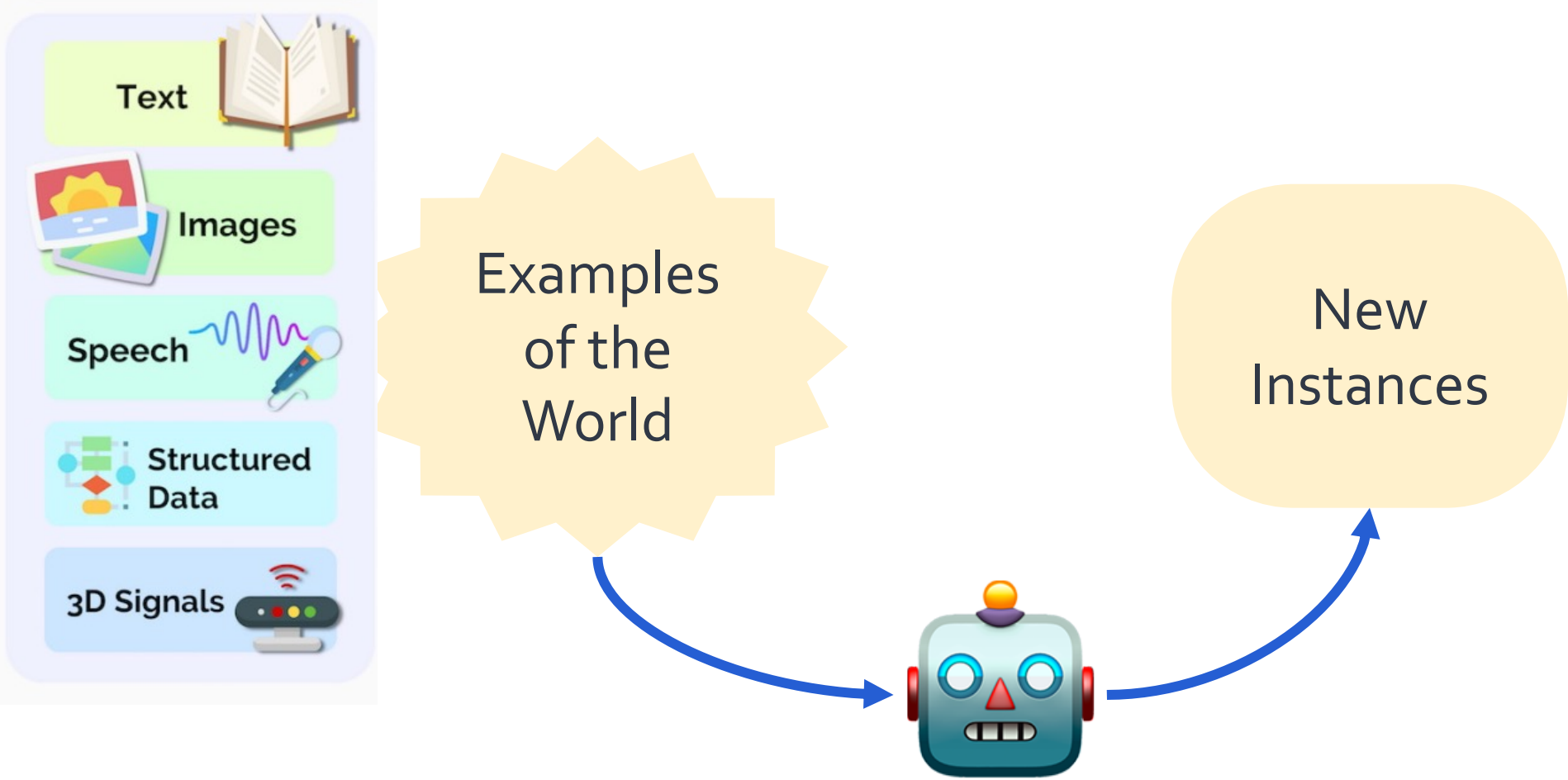
Are predictive models of the world.
Are learned from unlabeled ***data***

Examples of the World

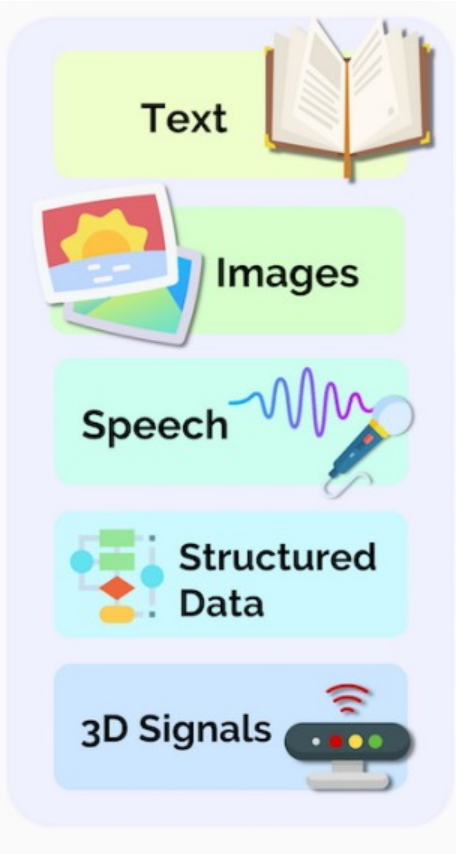


New Instances

Self-Supervised Models ~ Multi-modal Models



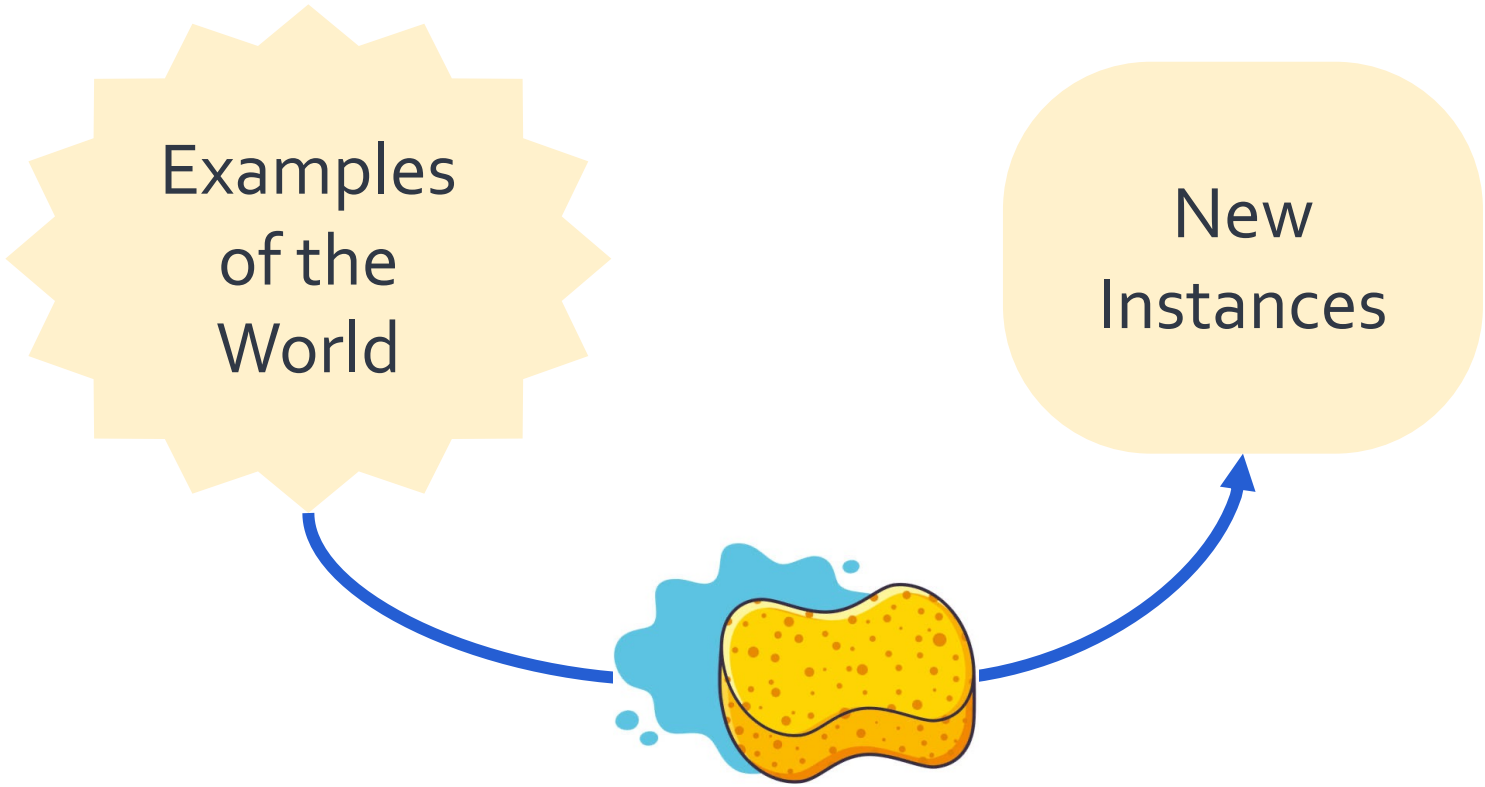
Generative AI Technologies



Examples of the World

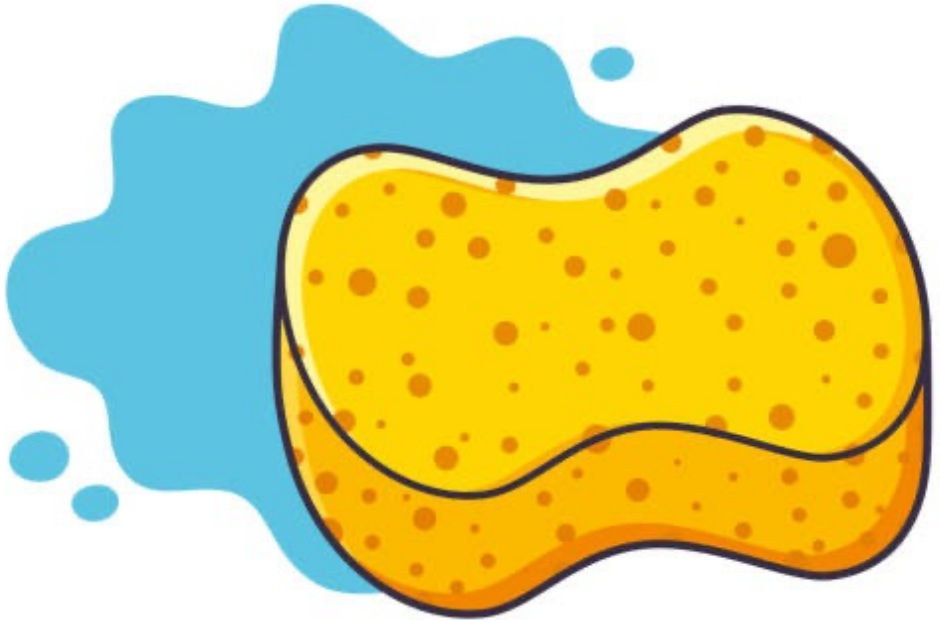
New Instances

AI Models ~ Sponges



AI Models ~ Sponges

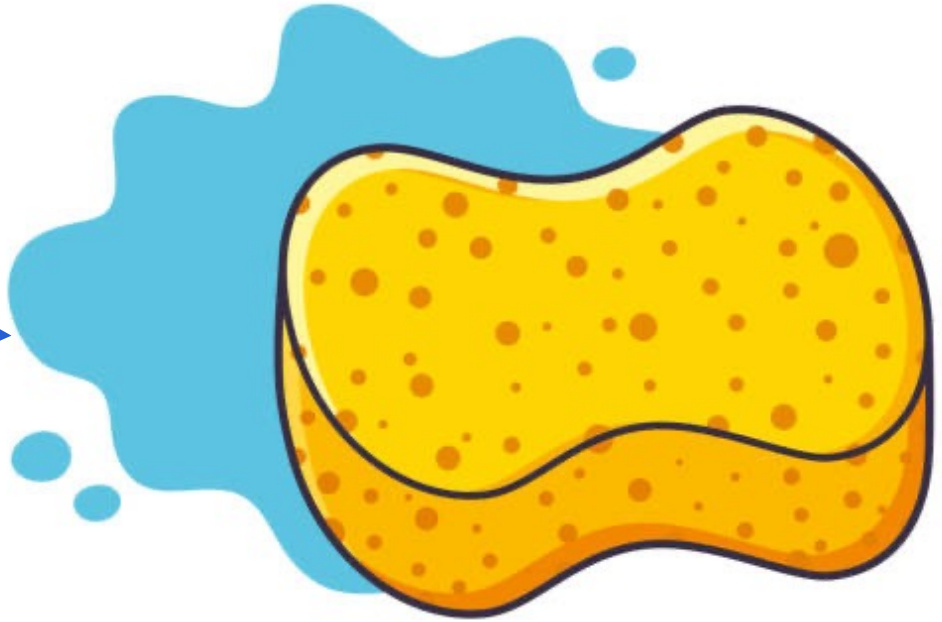
Examples of the World



New Instances

AI Models ~ Sponges

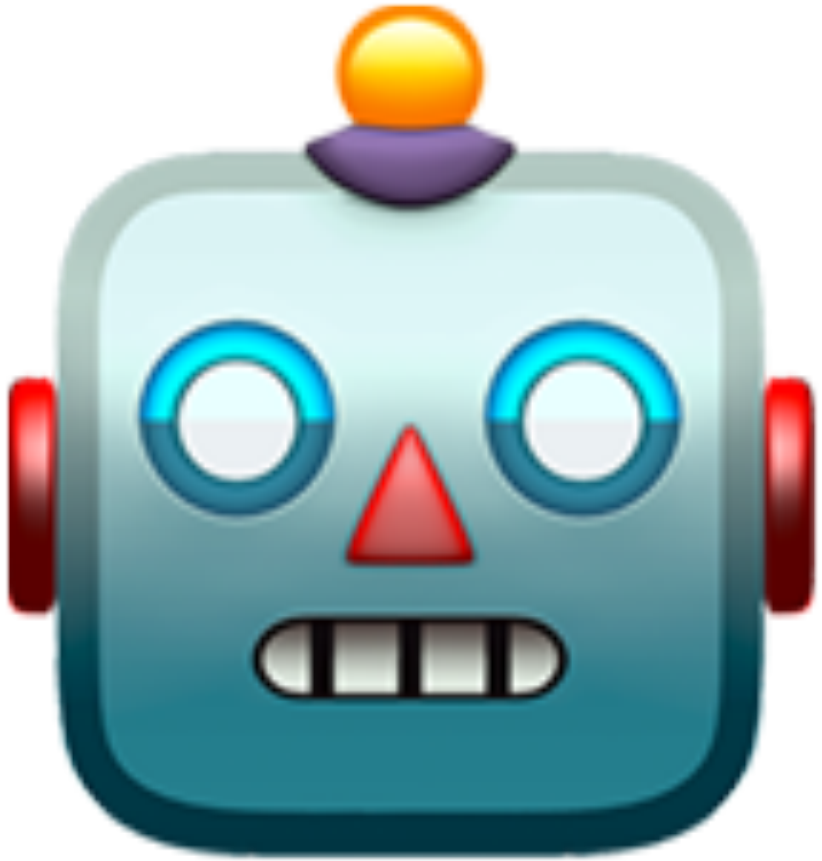
Examples of the World



New Instances

Scaling up AI Models Makes it Better

Examples of the World



New Instances

Scaling up AI Models Makes it Better

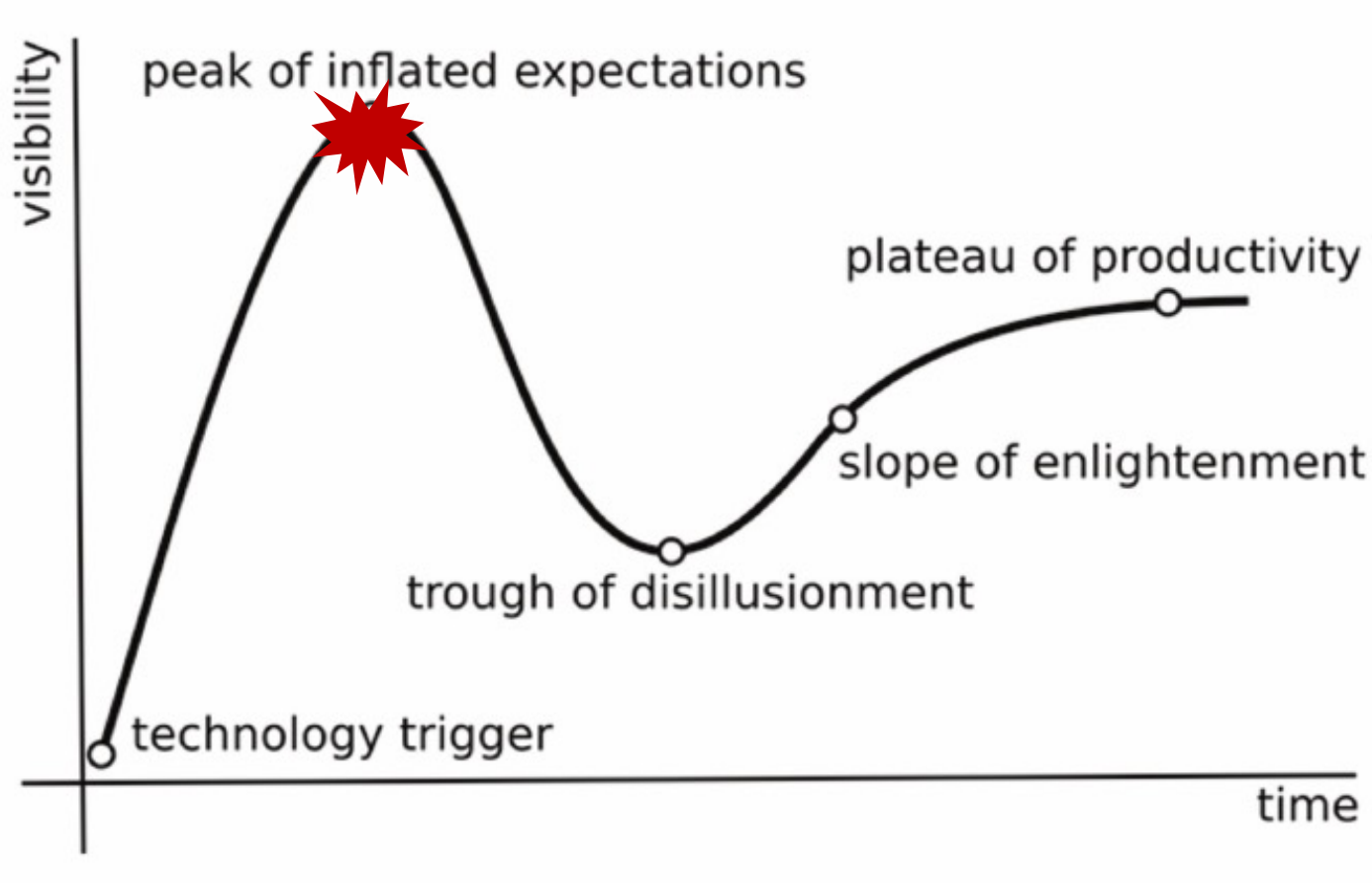
Why not just scale all the way down???

Gartner's
hype cycle

Scaling up AI Models Makes it Better

Why not just scale all the way down???

Gartner's hype cycle



Zipf's Law — The long tail
Most things are infrequent

- Nassim Nicholas Taleb suggests that biological & social dynamics lead to asymptotic distributions

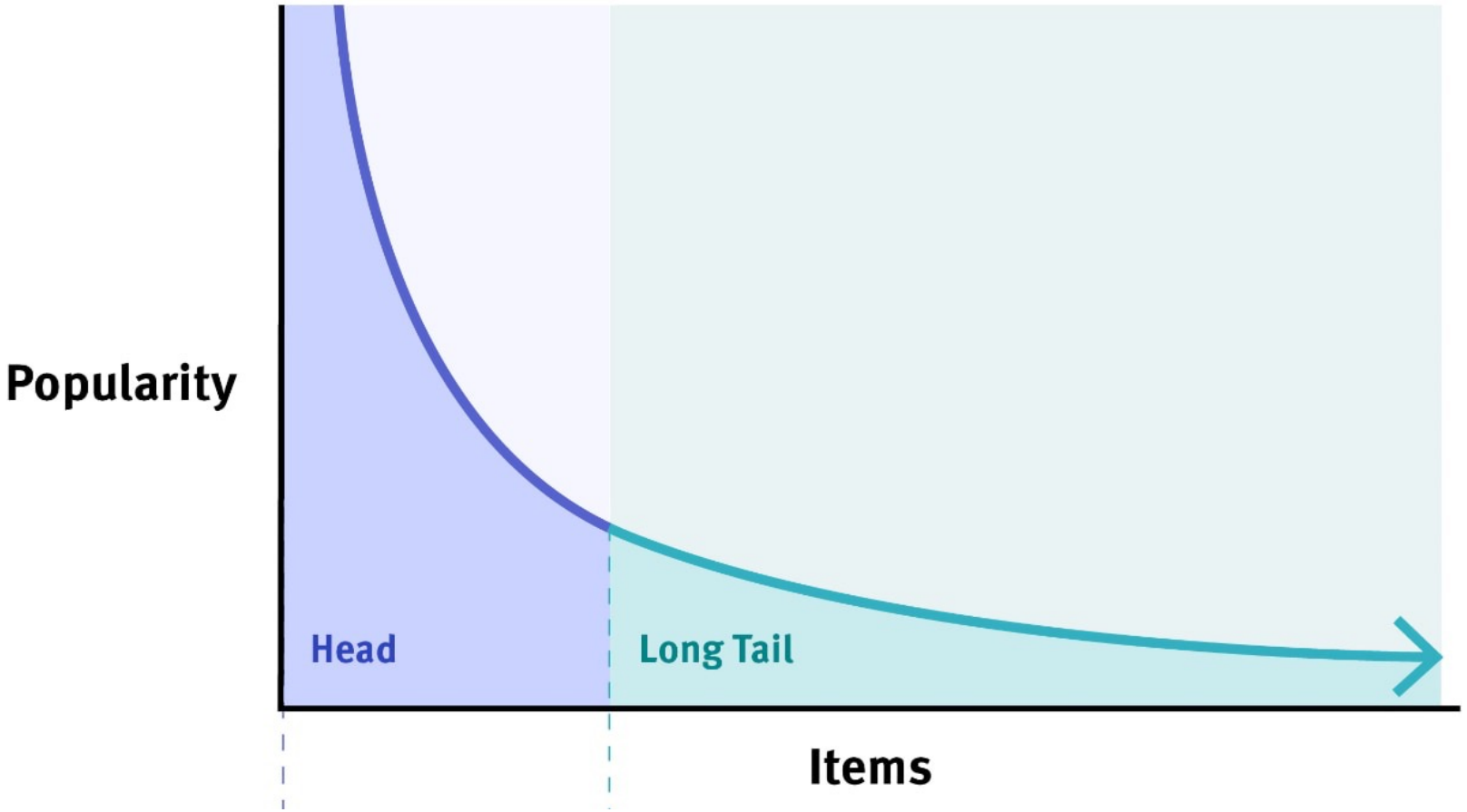
Zipf's Law — The long tail
Most things are infrequent

- Nassim Nicholas Taleb suggests that biological & social dynamics lead to asymptotic distributions

- **Examples:**
  - Wealth,
  - popularity,
  - number of sales of books,
  - number of views on social media,
  - frequency of a word,
  - many other social phenomena …

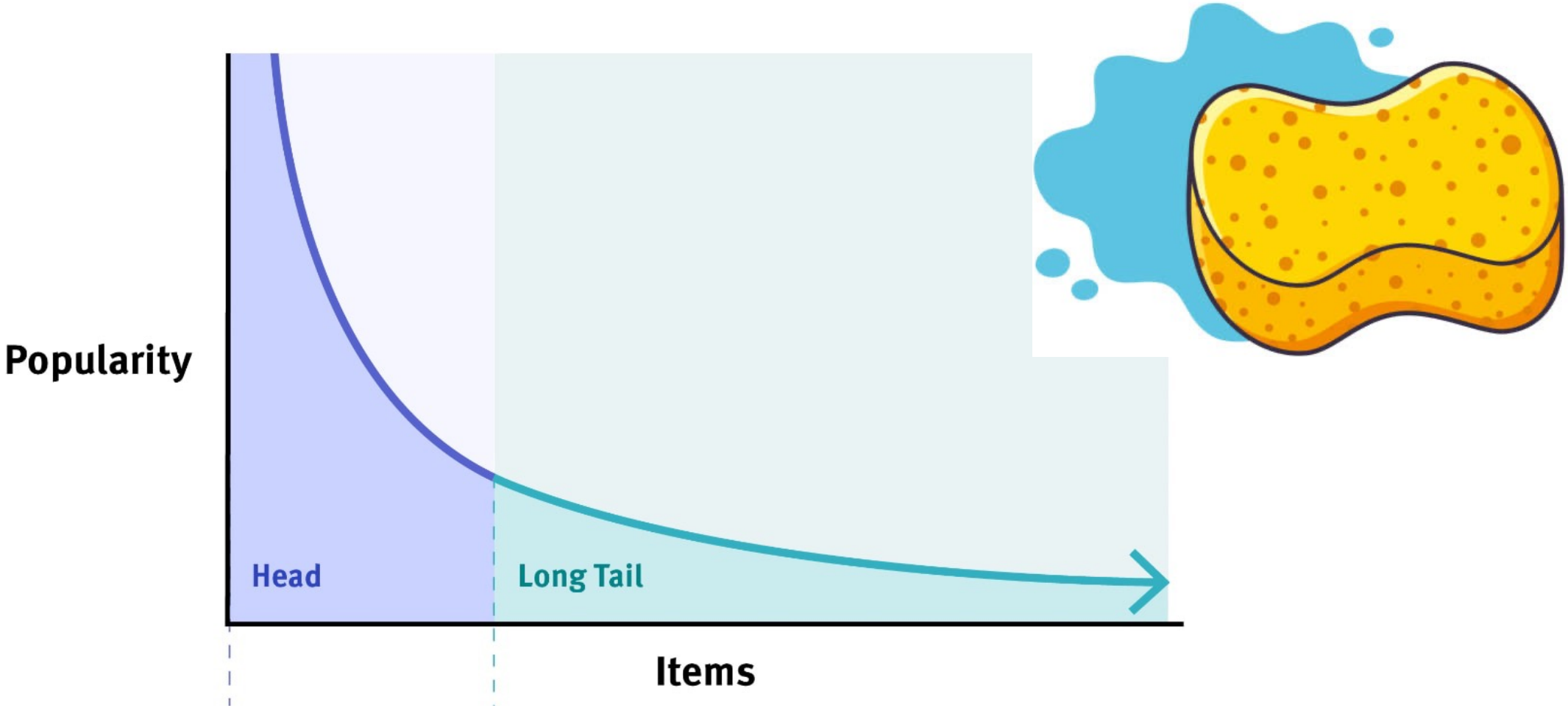
Zipf's Law — The long tail

Most things are infrequent



Zipf's Law — The long tail

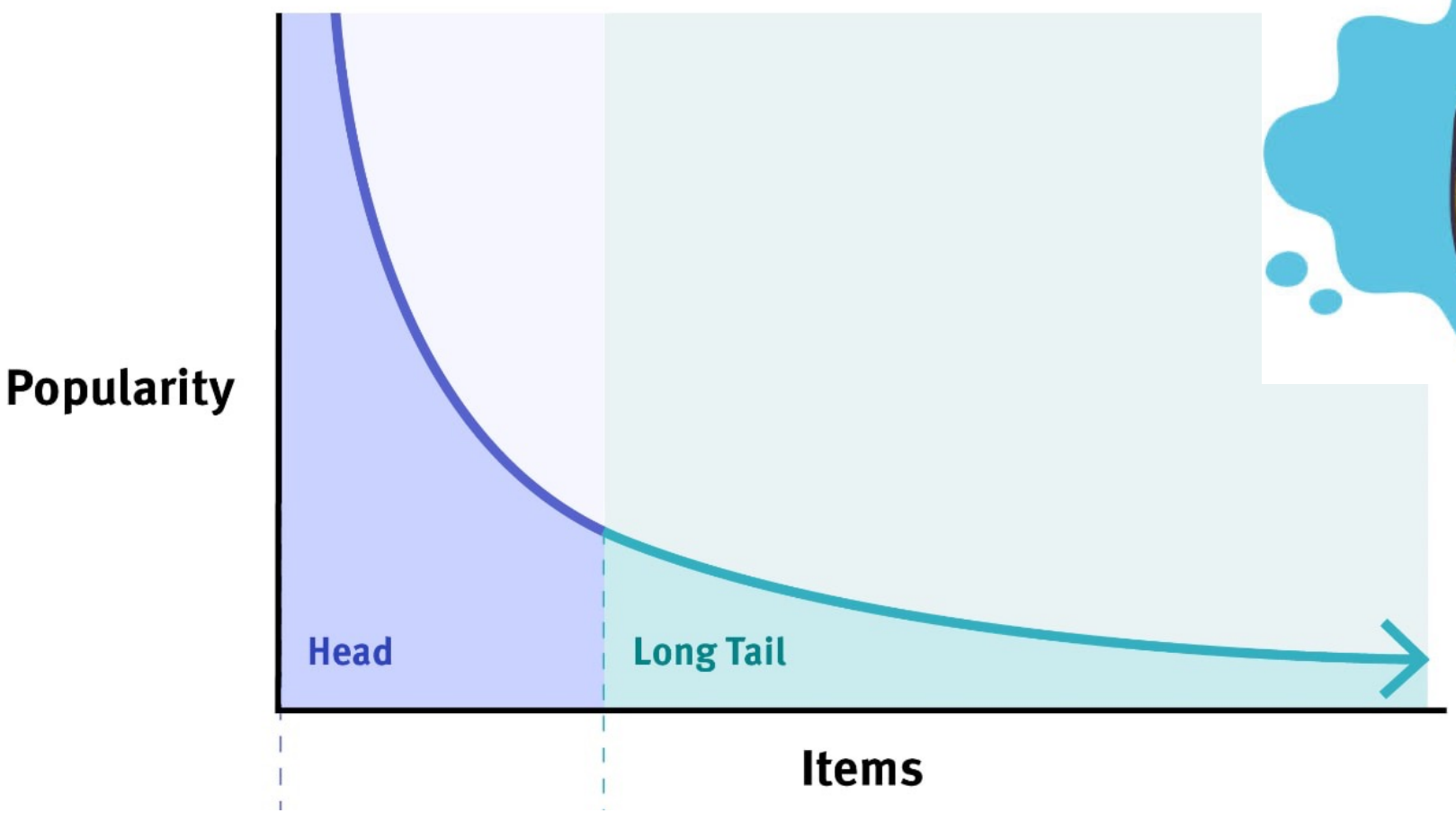
Models' quality depend on frequency of patterns

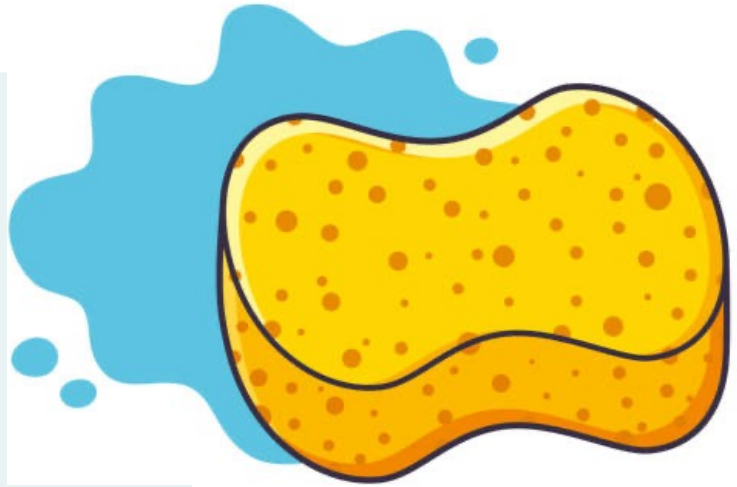


Zipf's Law — The long tail

Models' quality depend on frequency of patterns

The long tail => Hallucinations





Zipf's Law — The long tail

Models' quality depend on frequency of patterns

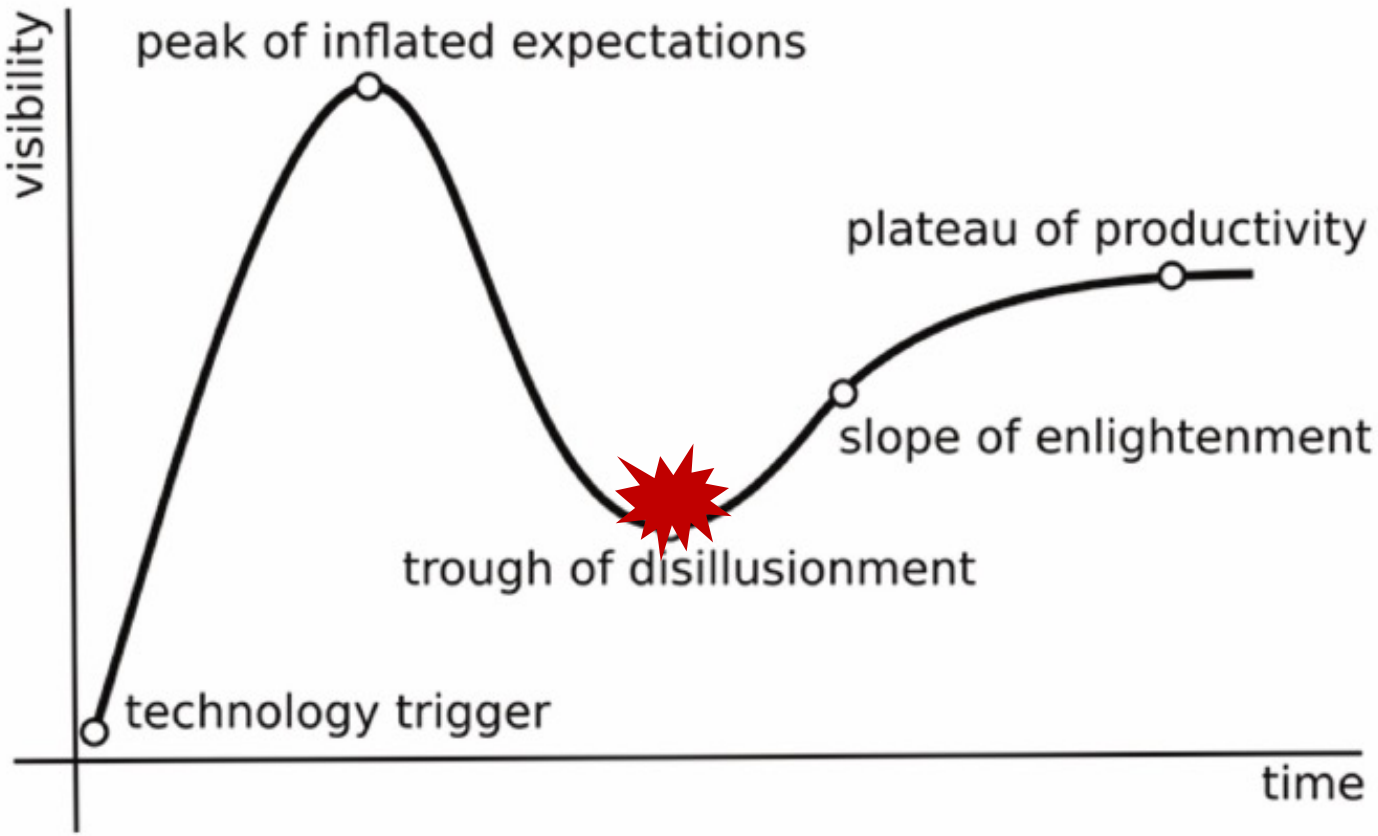
The long tail => Hallucinations

Zipf's Law — The long tail

Models' quality depend on frequency of patterns

The long tail => Hallucinations

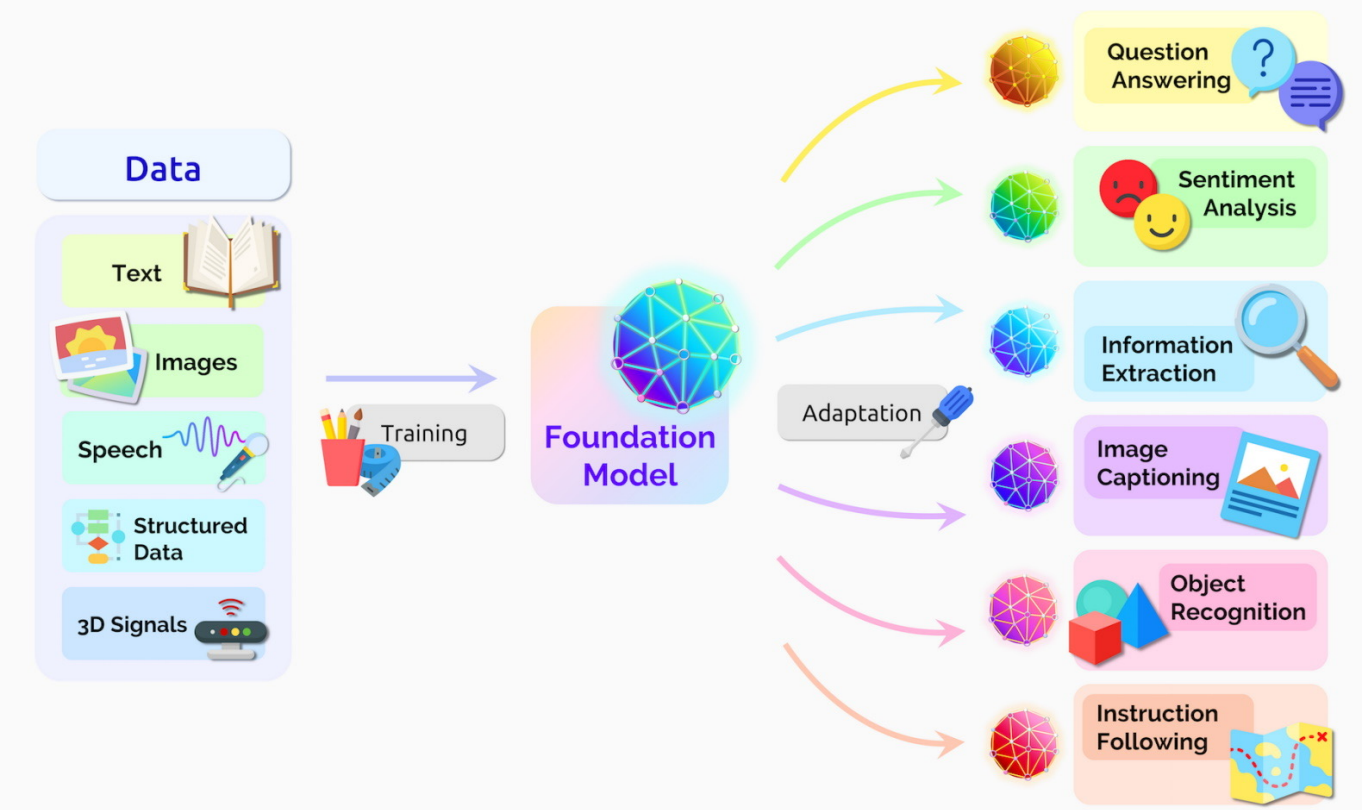
Gartner's hype cycle



Short Term: AI as a Tool

Specialized applications allow us to constrain the model to focused use cases and overcome hallucination.

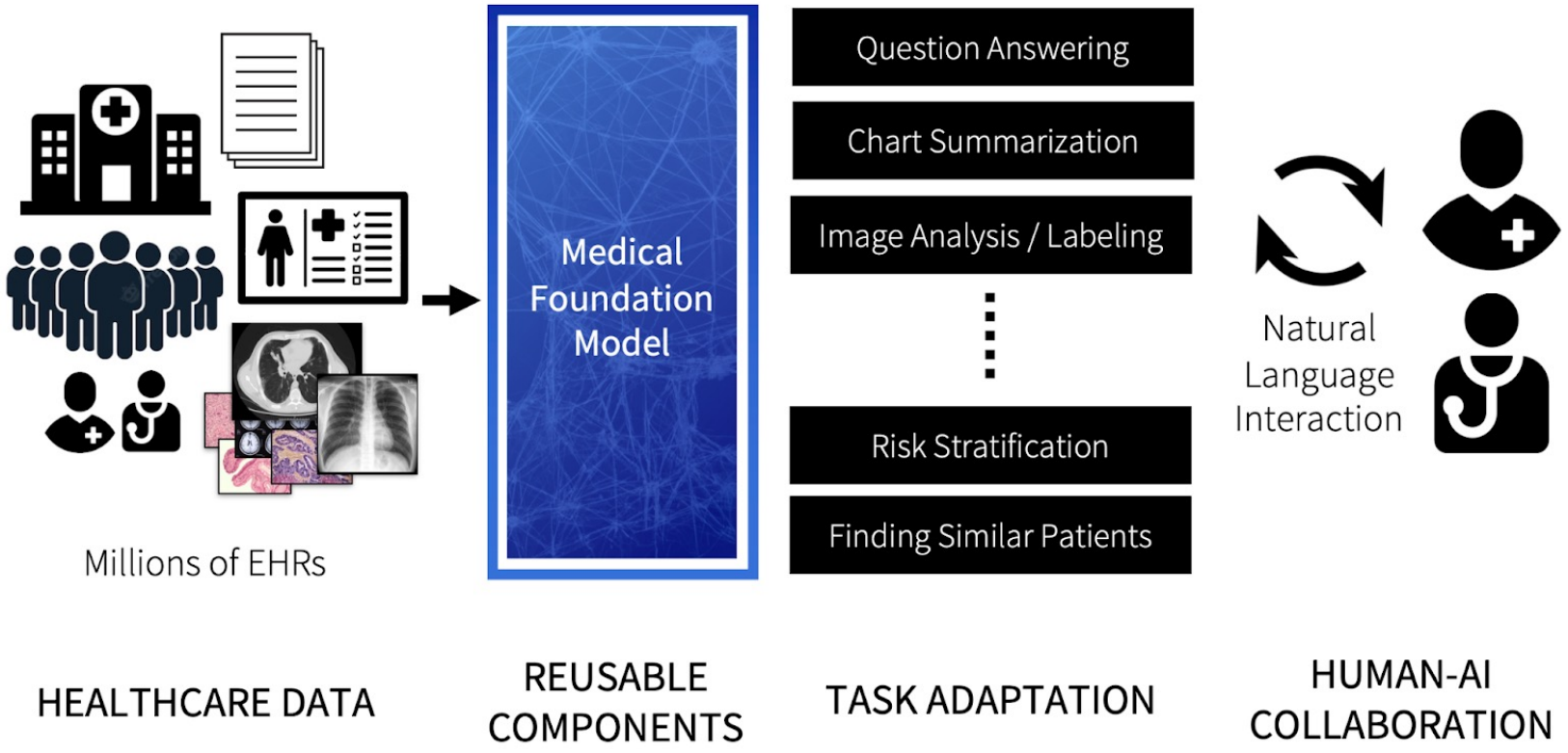
- Just like how calculator or microscope are tools.



Short Term: AI as a Tool

Specialized applications allow us to constrain the model to focused use cases and overcome hallucination.

- Similarly in medical applications



https://hai.stanford.edu/news/how-foundation-models-can-advance-ai-healthcare

Highlight: Information Needs in Healthcare

"Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery", March 2023

Highlight: Information Needs in Healthcare

- Can GPT-x can serve medical information needs in a safe and concordant manner?
- **Experiment:** Prompt models with common clinical questions:

  - *"93% of GPT-4 responses are deemed **safe**"*
  - *"41% of GPT-4 responses **agreed with the known answer**"*

"Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery", March 2023

Highlight: Information Needs in Healthcare

- Can GPT-x can serve medical information needs in a safe and concordant manner?
- **Experiment:** Prompt models with common clinical questions:

"In patients at least 18 years old who are prescribed ibuprofen, is there any difference in peak blood glucose after treatment compared to patients prescribed acetaminophen?"

- *"93% of GPT-4 responses are deemed **safe**"*
- *"41% of GPT-4 responses **agreed with the known answer**"*

"Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery", March 2023

Highlight: Information Needs in Healthcare

"Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum", 2023

Highlight: Information Needs in Healthcare

- Comparing responses of physicians vs. chatbot to patient questions
- **Experiment:** Questions posted to a public social media forum

- *".. evaluators preferred chatbot responses to physician responses in 78.6% of the evaluations."*

"Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum", 2023

Highlight: Information Needs in Healthcare

- Comparing responses of physicians vs. chatbot to patient questions
- **Experiment:** Questions posted to a public social media forum

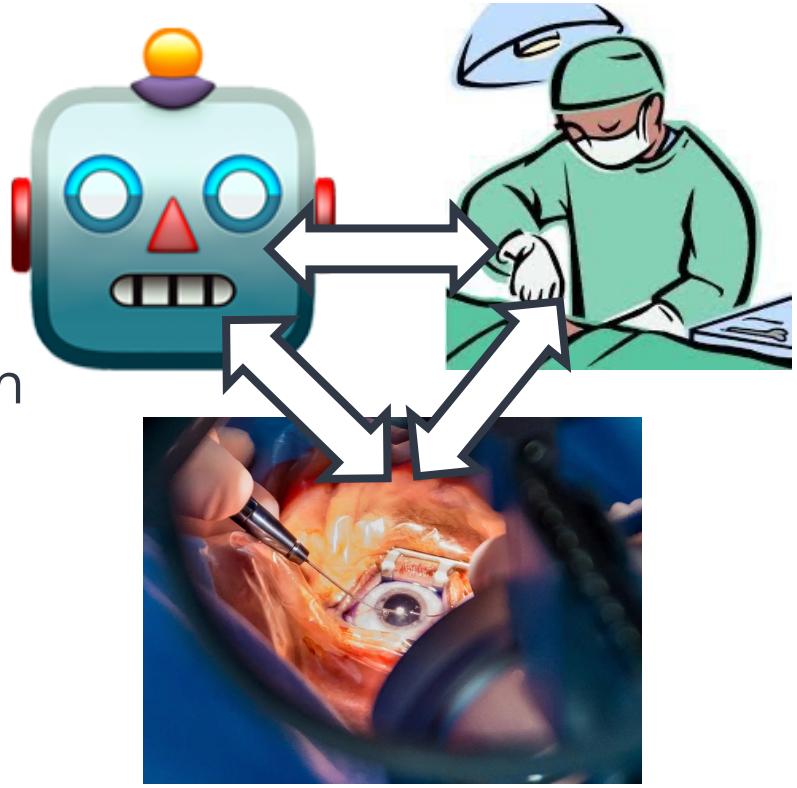
"Question about risk of dying following swallowing and ingesting a toothpick."

- *".. evaluators preferred chatbot responses to physician responses in 78.6% of the evaluations."*

"Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum", 2023

Highlight: Surgery Assistant

- **Goal:** an assistant to surgeon
  - Would provide on their weaknesses
  - Differences with other surgeons
  - Can also be used for bookkeeping in surgery room



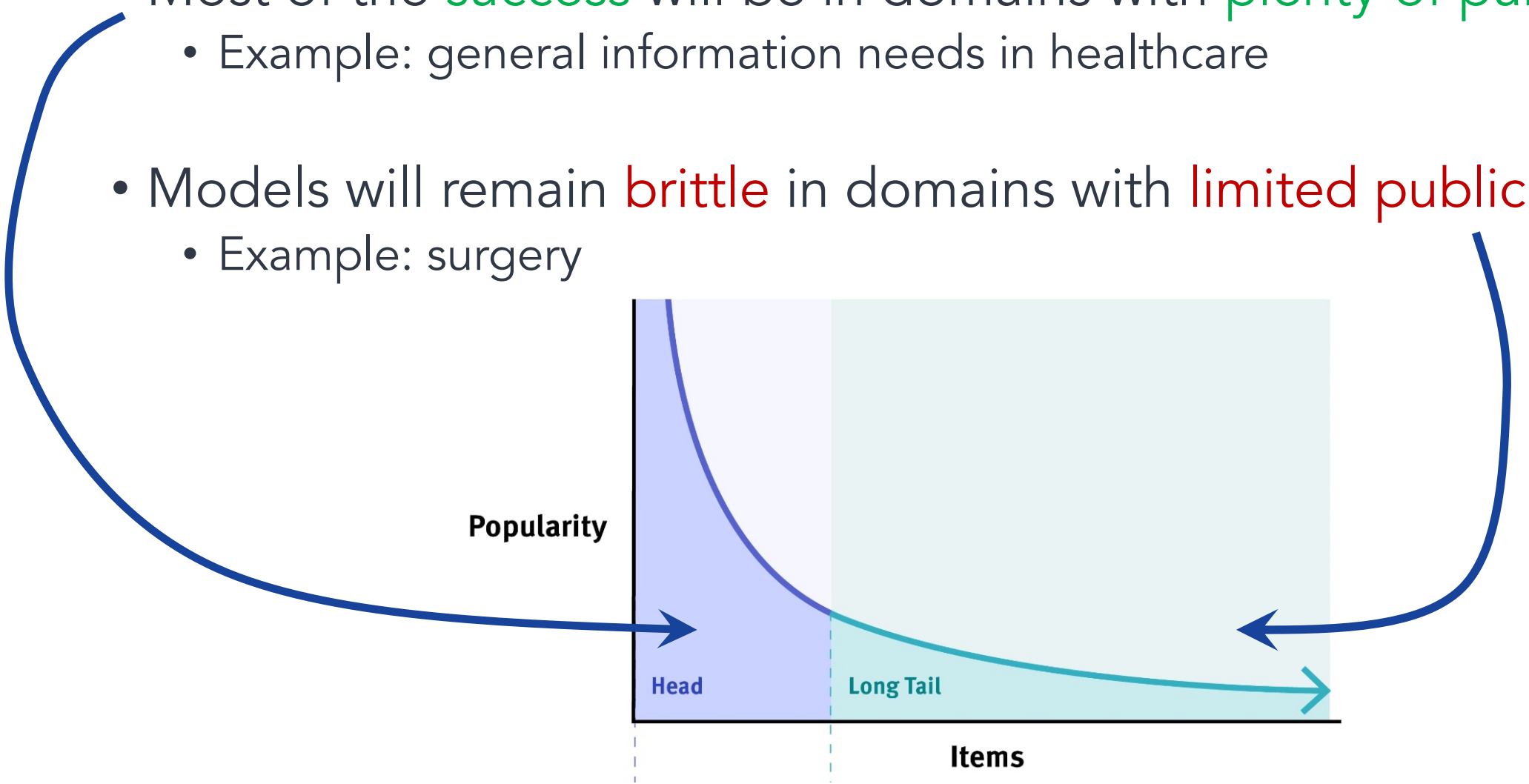
Our focus: eye surgery

- The problem is much more difficult due to limited data

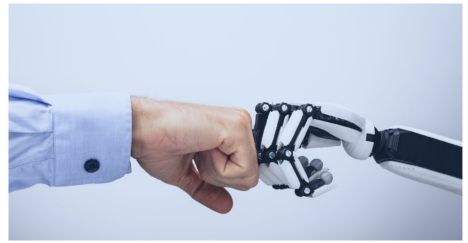
PIs: Swaroop Vedula, Shameema Sikder, Vishal Patel, Greg Hager, …

Going Forward …

- Most of the success will be in domains with plenty of public data.
  - Example: general information needs in healthcare
- Models will remain brittle in domains with limited public data.
  - Example: surgery

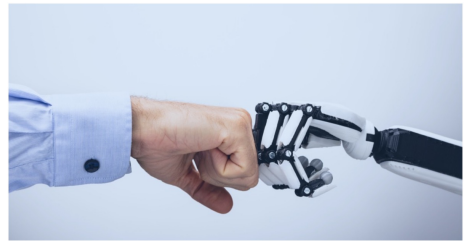


Human-AI Collaboration



Human-machine complementarity will bring opportunities

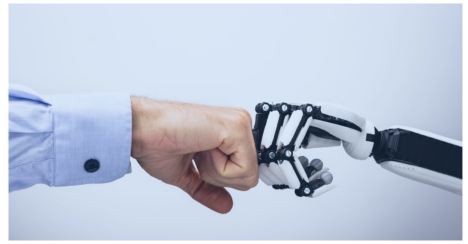
Human-AI Collaboration



Human-machine complementarity will bring opportunities

Things easy for humans

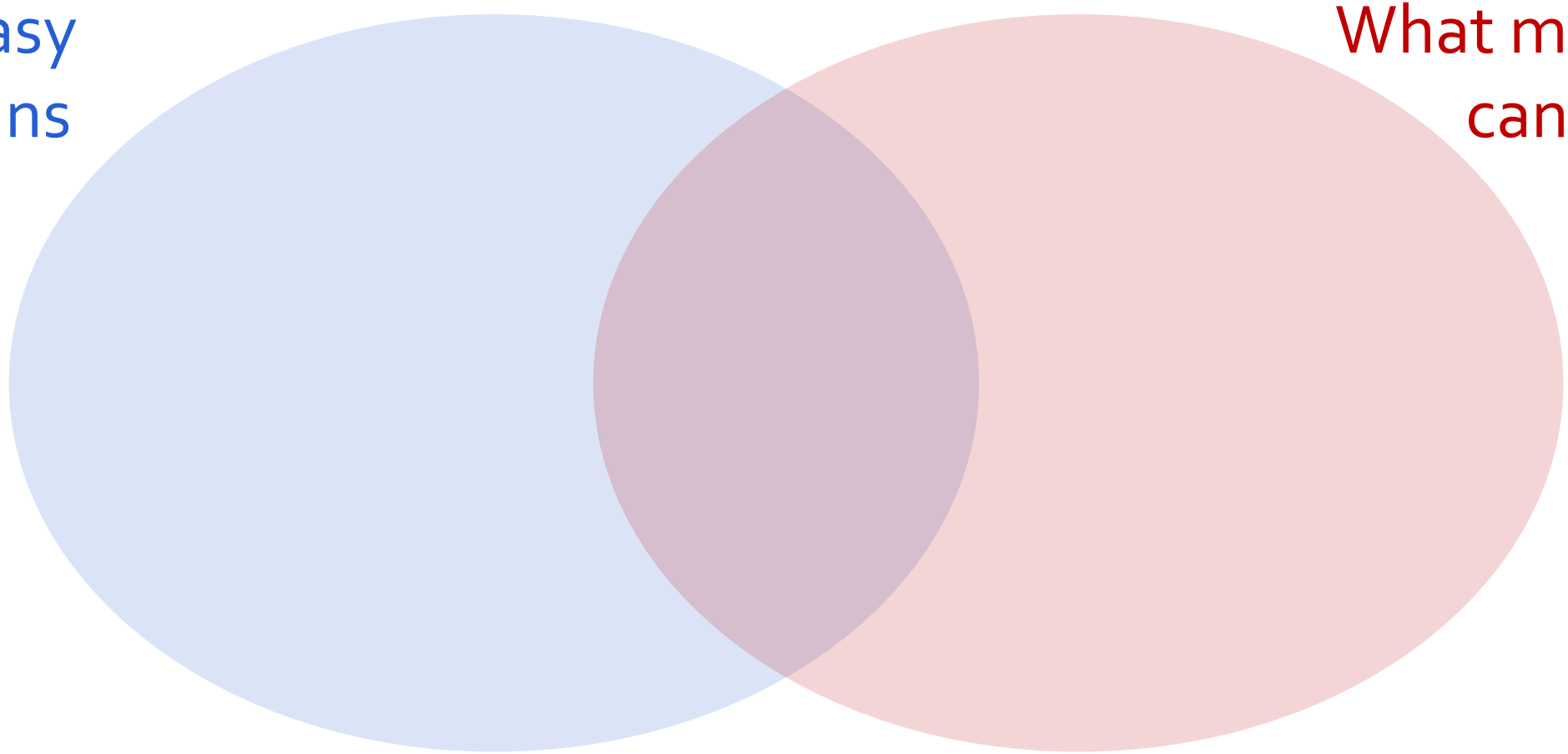
Human-AI Collaboration



Human-machine complementarity will bring opportunities

Things easy for humans

What machines can do



Human-AI Collaboration

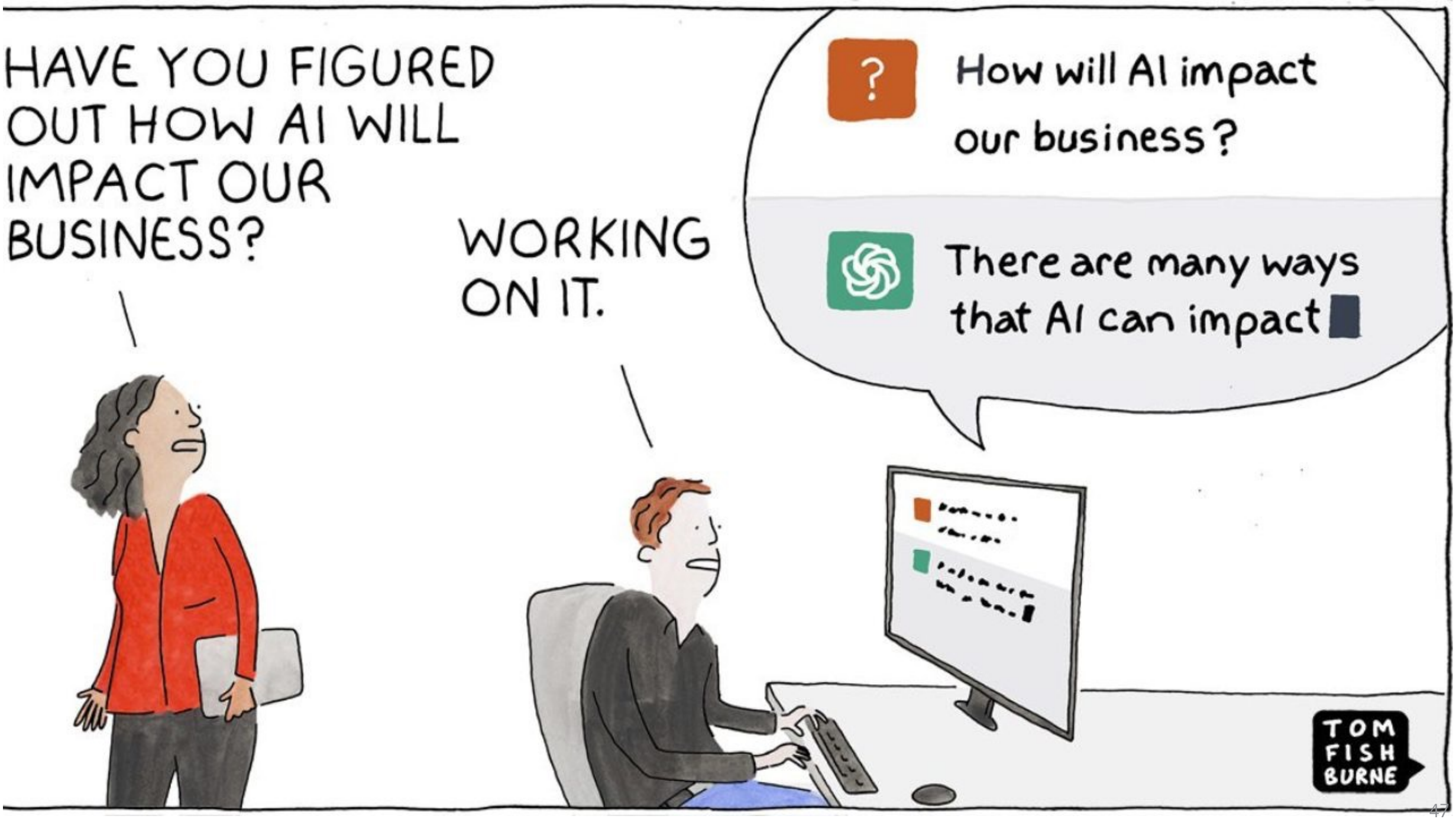
Human-machine complementarity will bring opportunities

Things easy for humans

What machines can do

Opportunity for human-AI colab

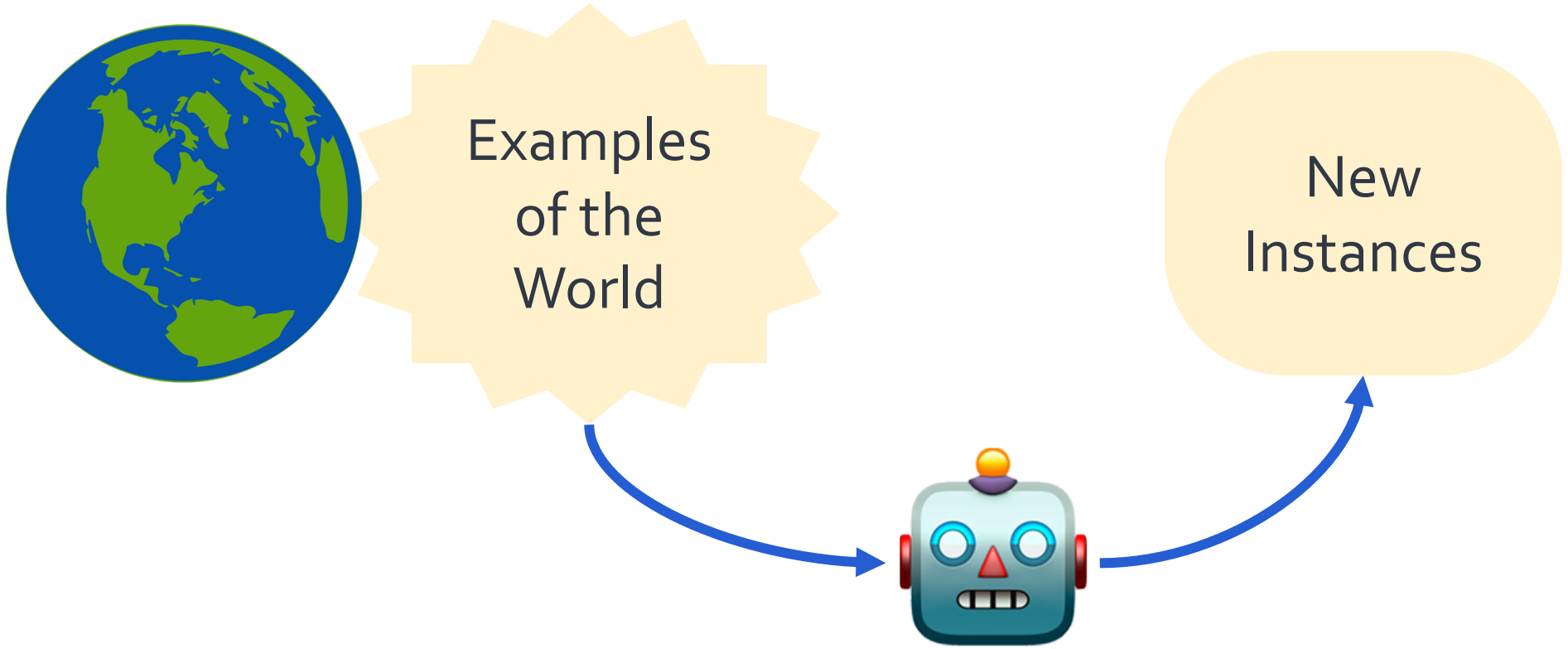
things that humans are slow at


HAVE YOU FIGURED OUT HOW AI WILL IMPACT OUR BUSINESS?
WORKING ON IT.
How will AI impact our business?
There are many ways that AI can impact
TOM FISH BURNE

Let’s talk about
concerns …

Danger: Web-Scale Poisoning

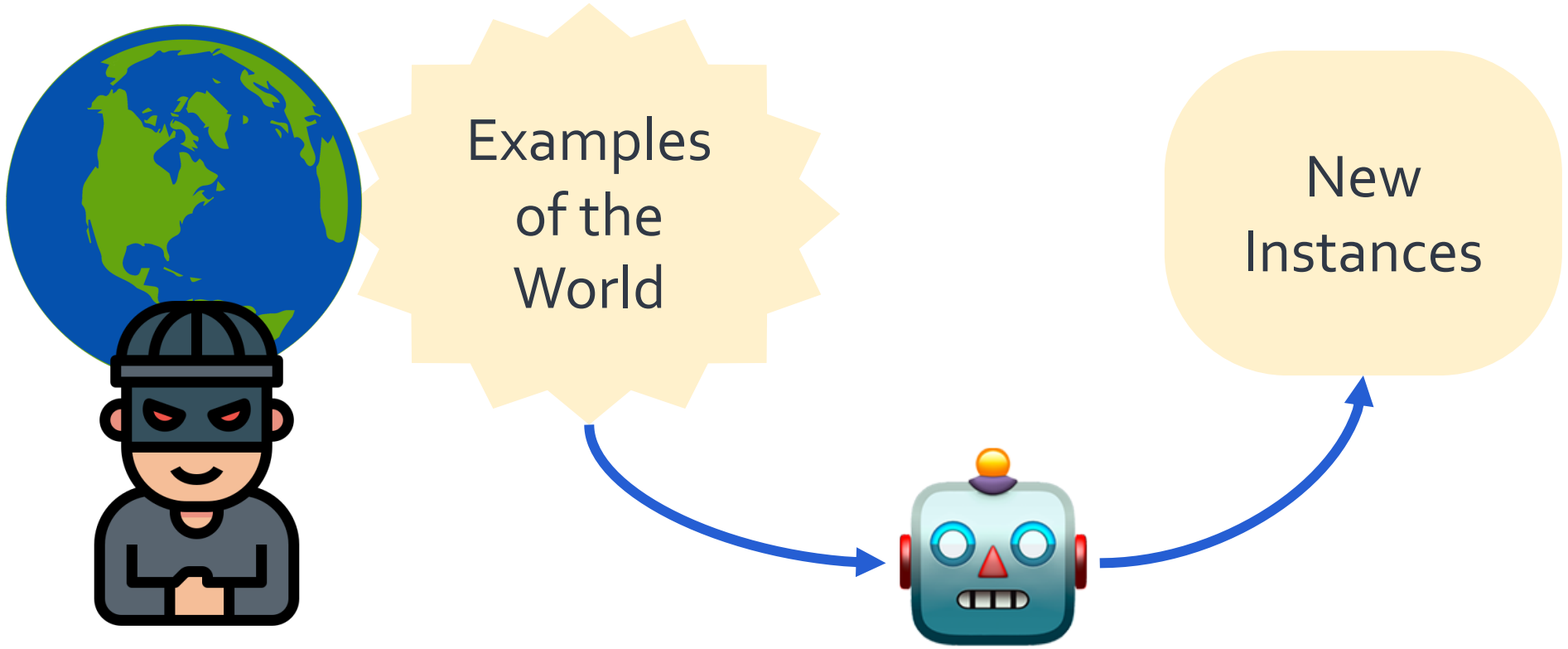
- Remember we said, most of the success will be in domains with plenty of public data.



"Poisoning Web-Scale Training Datasets is Practical", 2023

Danger: Web-Scale Poisoning

- Remember we said, most of the success will be in domains with plenty of public data.



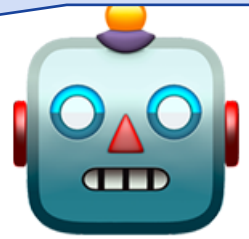
"Poisoning Web-Scale Training Datasets is Practical", 2023

Danger: Web-Scale Poisoning

- Remember we said, most of the success will be in domains with plenty of public data.

A wide range of options for "poisoning":

- Adding incorrect information
- Changing the distribution of the factual data
- Adding or modifying the URLs
- …



"Poisoning Web-Scale Training Datasets is Practical", 2023

Danger: Pushing Envelope of Privacy

Danger: Pushing Envelope of Privacy

- For models to function, they need more data.
- Collecting samples from individuals
  - Your emails, documents, excel sheets, medical records, biological or physical samples, TikTok pattern, etc.

Danger: Pushing Envelope of Privacy

- For models to function, they need more data.
- Collecting samples from individuals
  - Your emails, documents, excel sheets, medical records, biological or physical samples, TikTok pattern, etc.

- Age-old question, but now:
  1. More hunger for data
  2. Traces of data has become less conspicuous.

Danger: Pushing Envelope of Privacy

- For models to function, they need more data.
- Collecting samples from individuals
  - Your emails, documents, excel sheets, medical records, biological or physical samples, TikTok pattern, etc.

- Age-old question, but now:
  1. More hunger for data
  2. Traces of data has become less conspicuous.

- Easy answer: Ultimately the choice should be with the person who is giving up the data.
  - But do the users know what they're giving up?

Danger: Giving up Too Much + Sudden Change

- Gradually we are giving up more of our autonomy.
- There is a future where AI systems will be making more society-level decisions by optimizing various proxy objectives.
- On most days, things will work just fine.
- From time-to-time AI systems may fail catastrophically
  - Various society-level distributional drifts, such as a conflict between states, a natural disaster, a serious cyberattack, etc.

"What failure looks like", Paul Christiano 2019

Putting it Together

- I am excited and nervous.

Putting it Together

- I am excited and nervous.

- The opportunity to make many jobs more efficient is exciting.

- With expansion of digital society, there is increasing room for malicious activities in cyber space.

Other Concerns I Didn’t Cover

Other Concerns I Didn't Cover

- Copyright policy
  - My take: things won't change unless something bad happens.

- Over-optimizing for market.
  - Example: snapchat AI
  - Quite concerning.

Other Concerns I Didn't Cover

- Copyright policy
  - My take: things won't change unless something bad happens.
- Over-optimizing for market.
  - Example: snapchat AI
  - Quite concerning.
- Job market:
  - My take: real concern, but I am optimistic.
  - Change will be slower than people think.
- Concerns about super-intelligence
  - I am concerned.

Room for Collaborations

- It is likely that future regulations will involve algorithmic tests.
  - For example, test of whether a model has been exposed to certain copyrighted material.

Room for Collaborations

- It is likely that future regulations will involve algorithmic tests.
  - For example, test of whether a model has been exposed to certain copyrighted material.

- We understand the technical tools and challenges.
- You understand the policy/application landscape.
- Dream team? 😉

Thanks!