Foundation Models & ChatGPT :

A new territory in need of new maps

Roald Sieberath

LeanSquare | UCLouvain

TRAIL + VAIA seminar

Plan

- Intro : (+) BIASES
- WHY this is important
- WHAT are we talking about ?
  - reasons to be excited
  - reasons to be worried
  - questions to pursue
- HOW can we move forward ?
  - 3 proposals
    (+ surprise)

Intro : my *biases*

Machine Learning

by Andrew Ng

Machine Learning

coursera

Artificial Intelligence
Nanodegree Program

HAI

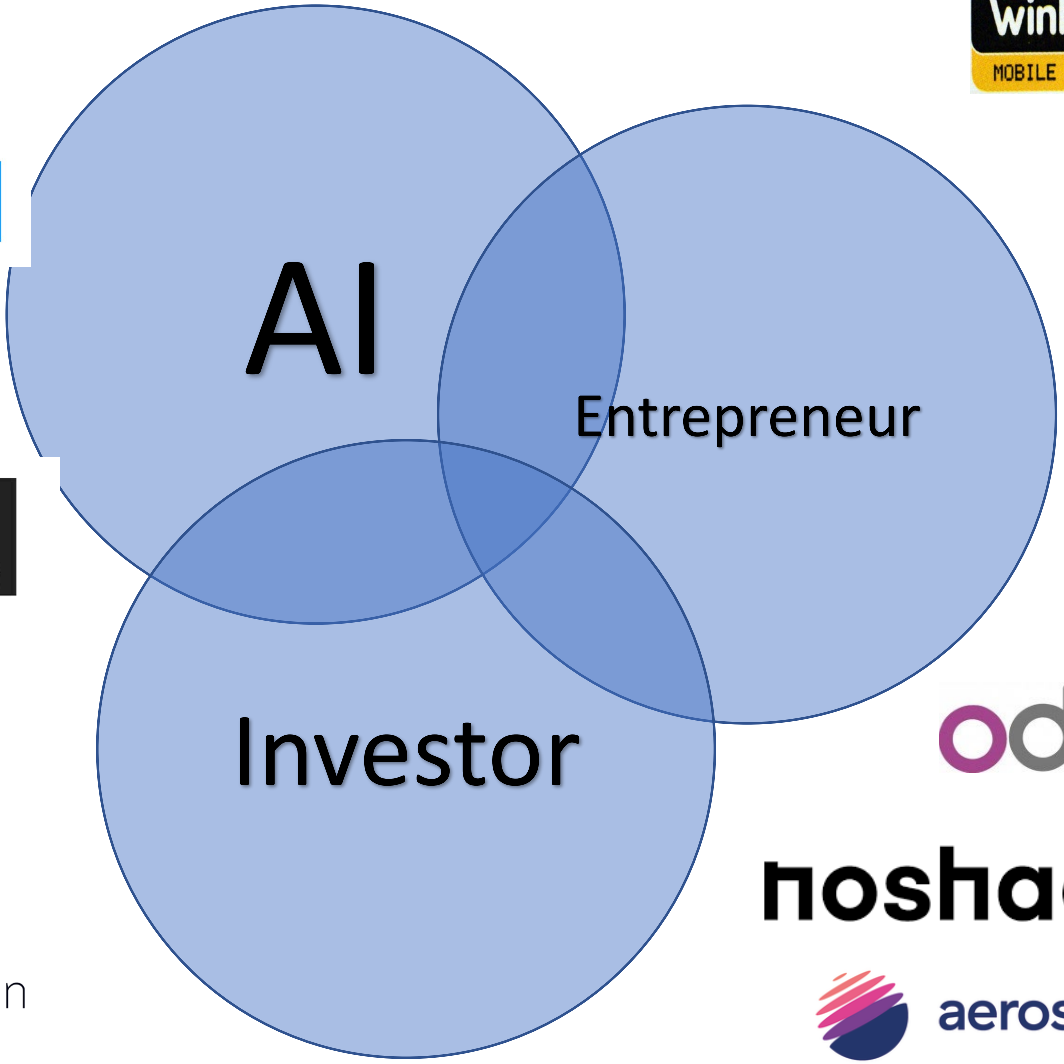
Stanford University
Human-Centered
Artificial Intelligence

fast.ai

Disclaimer:

Doubt everything I will be saying about AI ;-)

Kauffman
Fellows



winbox.com
MOBILE MESSAGING

SWAN INSIGHTS

AI
Black Belt

BE CENTRAL

odoo

noshaq

lean square

aerospacelab

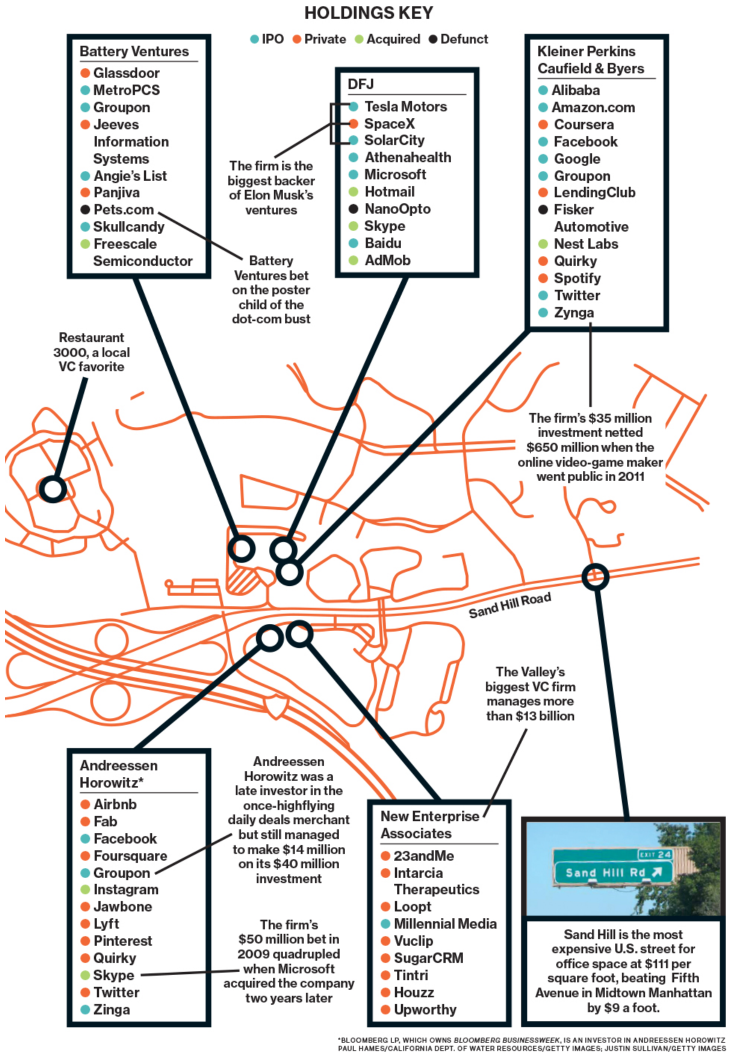

GLOBETROTTER
GLOBETROTTER
Academia
Business

Nexus of my biases : Silicon Valley

Stanford University
Human-Centered
Artificial Intelligence

Kauffman
Fellows

Global
Valley



Just one exemple

- **Lucas Biewald**
- TA in AI at Stanford (mid-2000)
- Saw the opportunity to build training datasets for supervised learning
- Started Crowdflower (clients : FB, Google…)
- Sold to Appen for 300 M$
- Saw the opportunity for MLops tools
- Started Weights & Biases, raised 200 M$

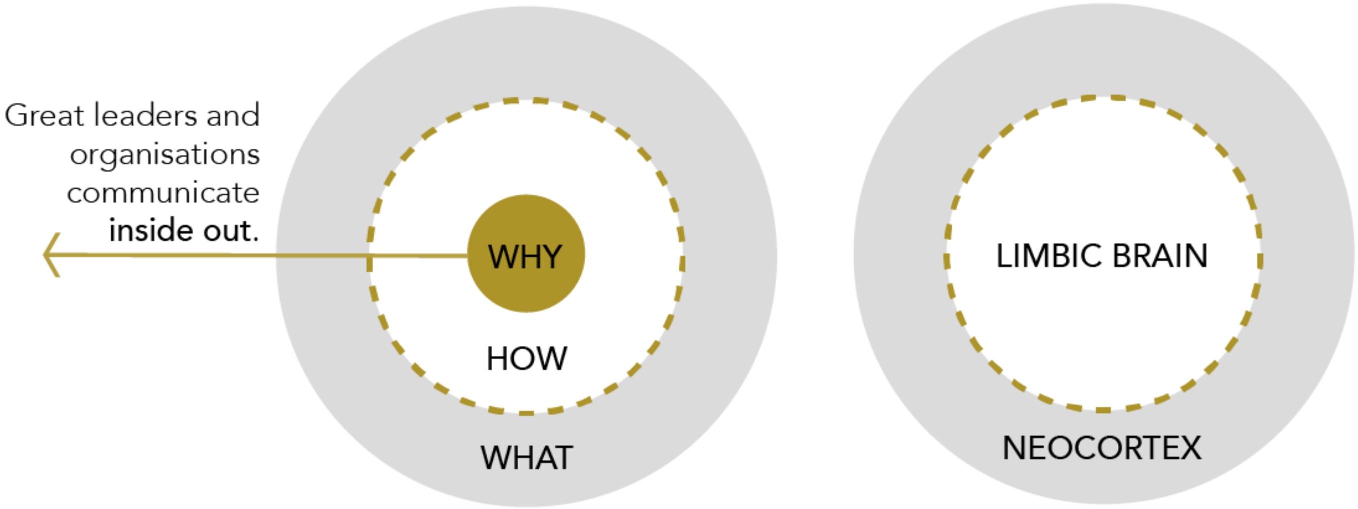
- Note : success in startups is not just IQ

CrowdFlower

Weights & Biases

WHY this is important?

The Golden Circle + Human Brain



Why - Your Purpose
Your motivation? What do you believe?

How - Your Process
Specific actions taken to realise your Why

Limbric Brain - Your Trust
Controls behavior and decision making
Result: 'Gut' feelings and loyalty

What - Your Result
What do you do? The result of Why - Proof

Neocortex - Your Rational
Controls senses, spatial reasoning, analytical thinking and language
Result: Rationalisation and communication

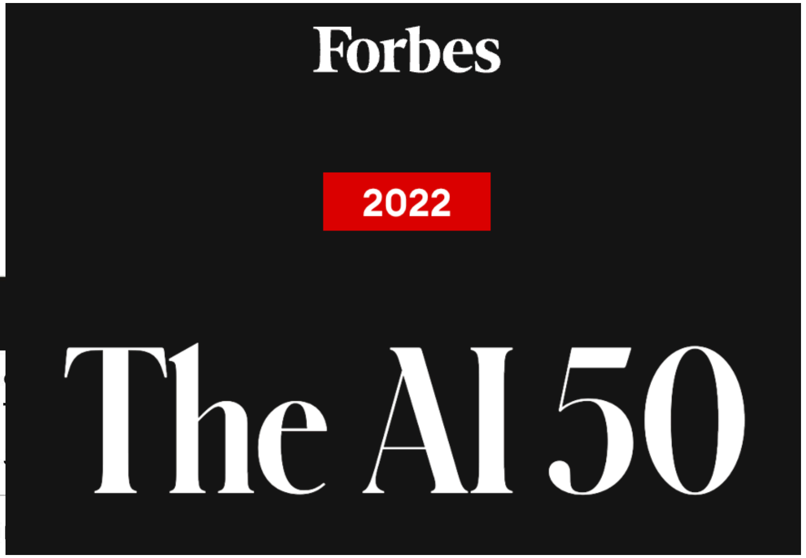
Simon Sinek, "Start with Why"

WHY this is important?

- European independence ?
  - Smartphones ?
  - Large software companies ?
  - Rockets ?
  - GPS –> Galileo
  - Starlink –> IRIS2

- AI ?

WHY this is important?



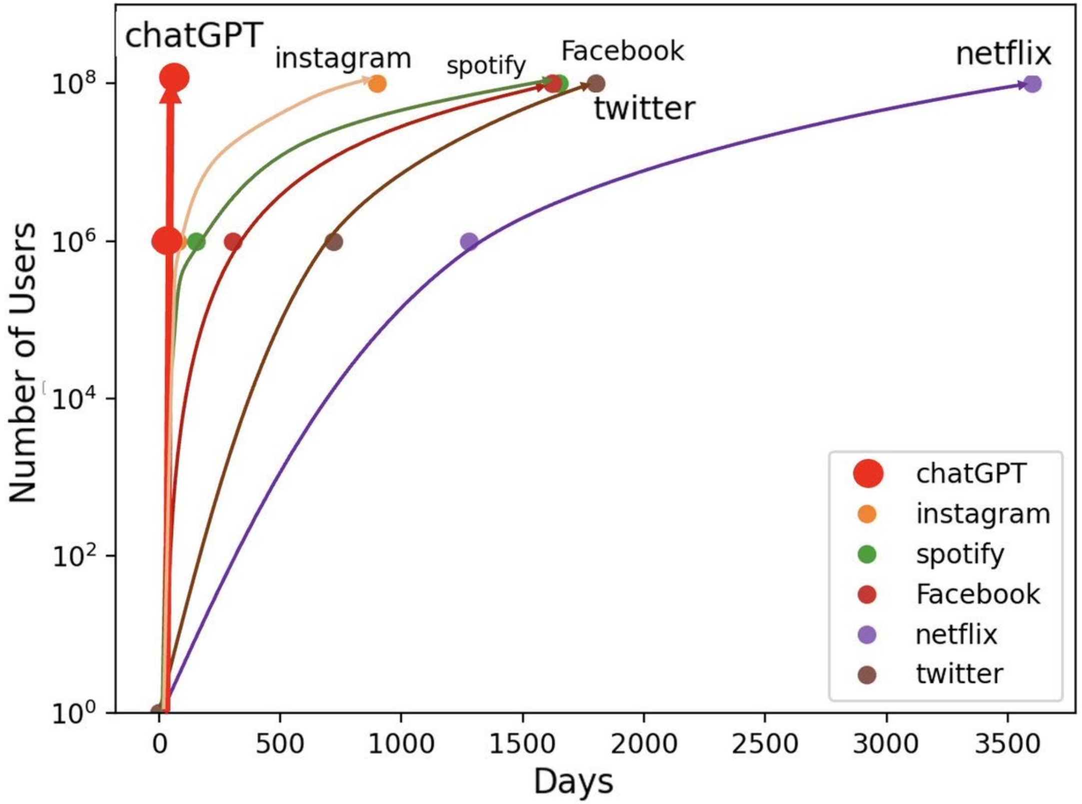
Forbes

NAME	INDUSTRY	FUNDING	HEADQUARTERS	
6sense	Sales and Marketing	$426 M	San Francisco, California, United States	
Abacus.AI	Data Science	$90 M	San Francisco, California, United States	
Abnormal Security	Cybersecurity	$74 M	San Francisco, California, United States	Evan Reiser
Amira Learning	Education	$21 M	San Francisco, California, United States	Mark Angel
AMP Robotics	Environment and Energy	$78 M	Louisville, Colorado, United States	Matanya Horowitz
Anyscale	AI Infrastructure	$160 M	San Francisco, California, United States	Robert Nishihara
Arize AI	Data Science	$23 M	Berkeley, California, United States	Jason Lopatecki
ASAPP	Customer Service	$400 M	New York, New York, United States	Gustavo Sapoznik
Aurora Solar	Environment and Energy	$523 M	San Francisco, California, United States	Christopher Hopper
Brain Technologies	Consumer Technology	$50 M	San Mateo, California, United States	Jerry Yue
Brightseed	Pharmaceutical	$115 M	San Francisco, California, United States	Jim Flatt
Canvas	Construction	$83 M	San Francisco, California, United States	Kevin Albert
ClosedLoop	Healthcare	$45 M	Austin, Texas, United States	Andrew Eye

Investors don’t look much at citation index

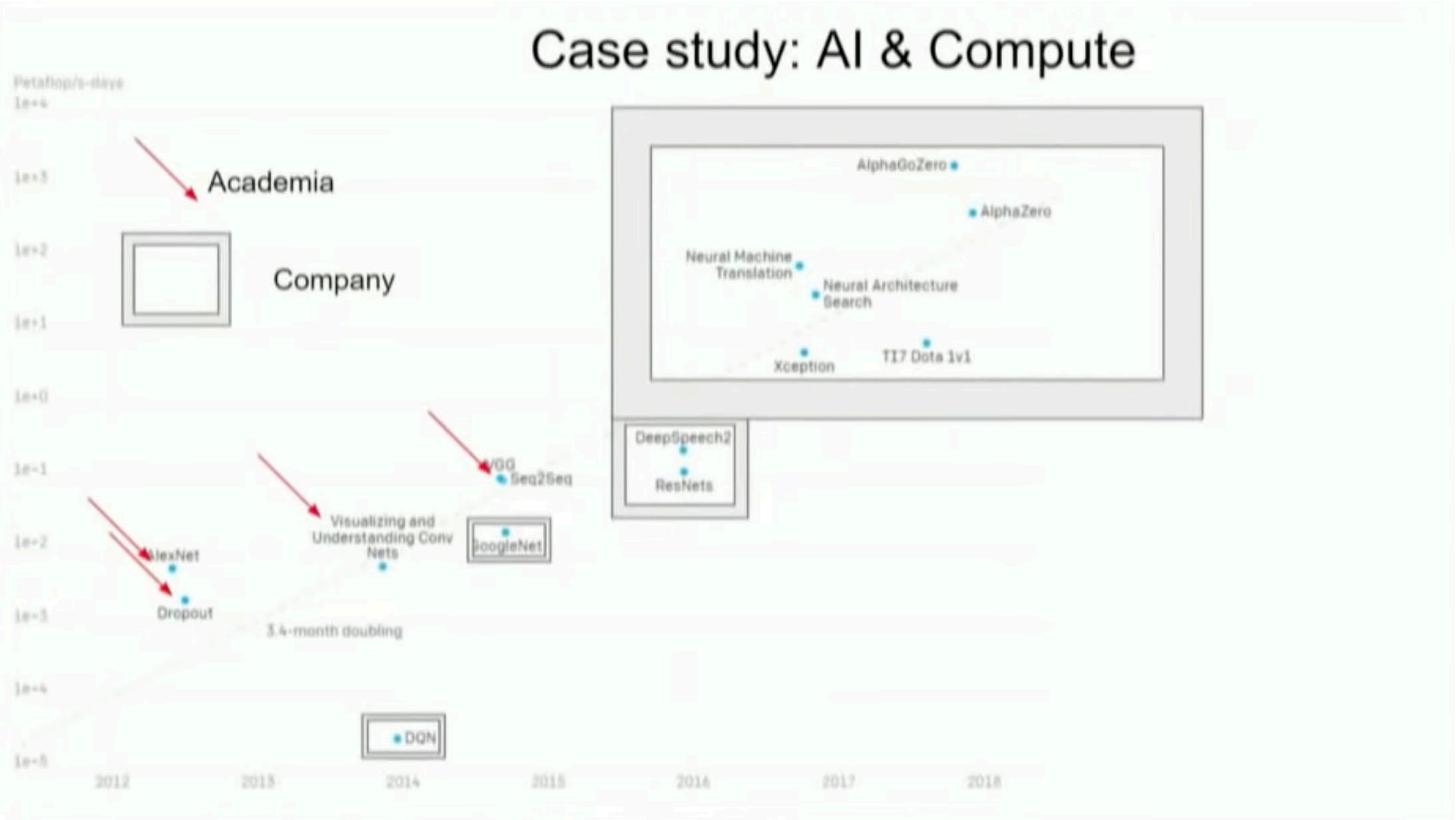
WHY this is important?

- Sense of urgency :
  - This is not “exponential” growth, this is *vertical*



WHY this is important?

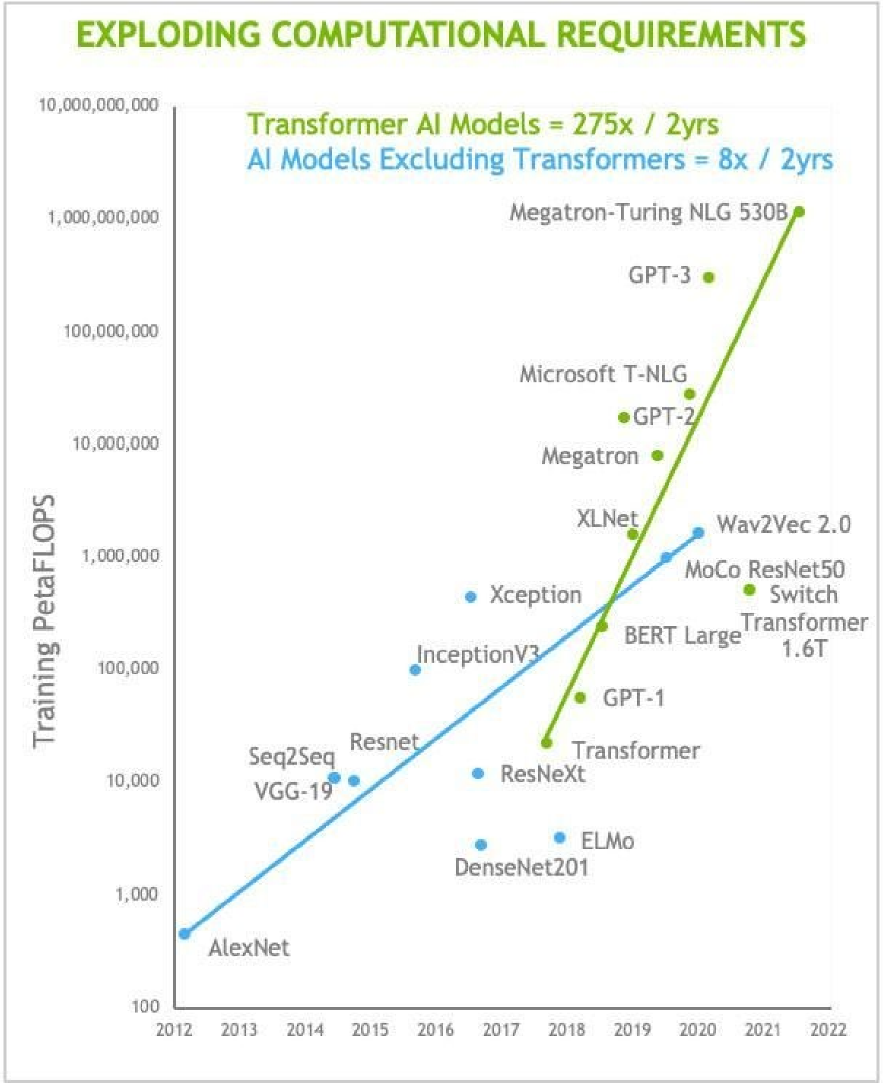
Compute resources needed go out of hand



“AI & Compute”, OpenAI blog, quoted by Jack Clark, Anthropic AI

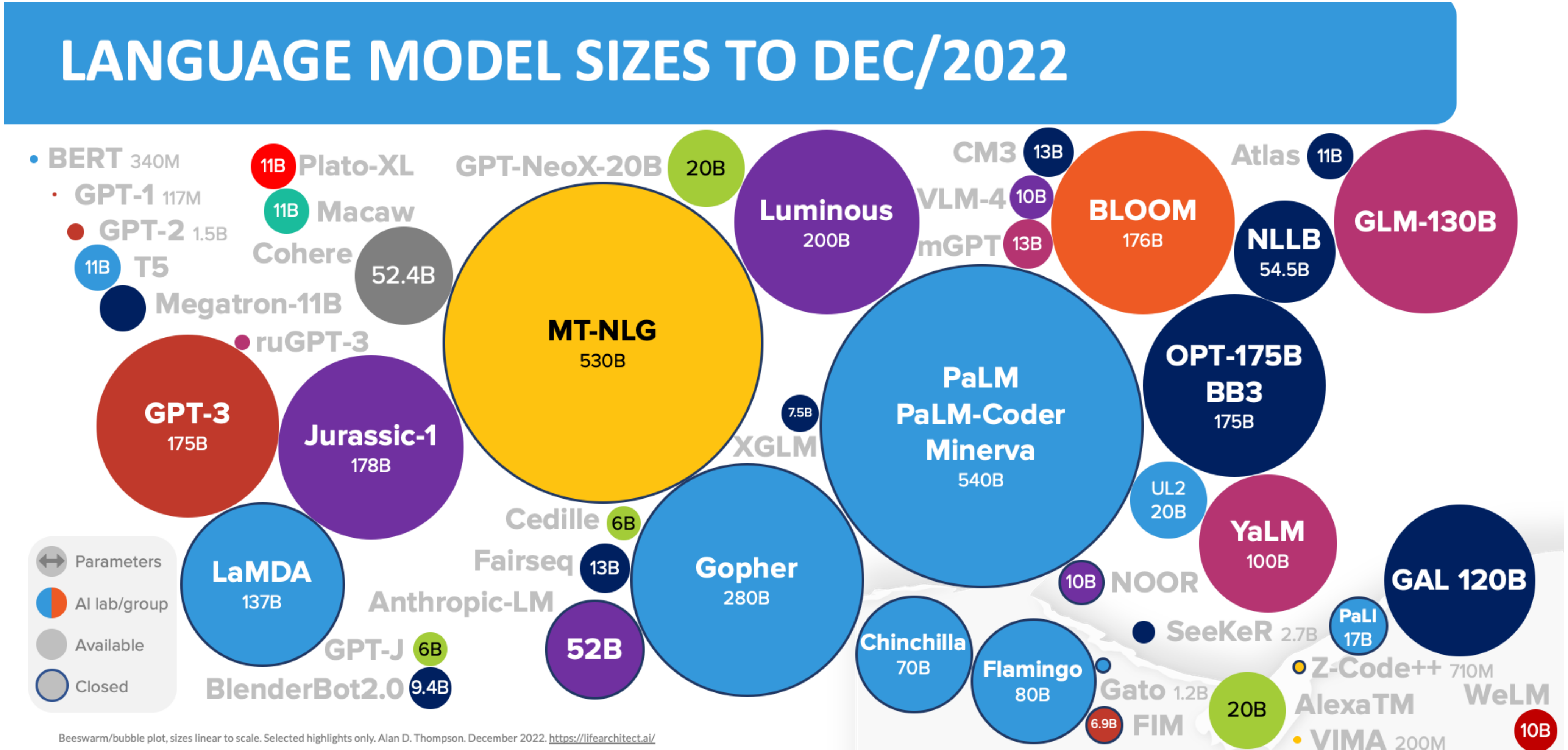
WHY this is important?

Compute resources
needed
go out of hand



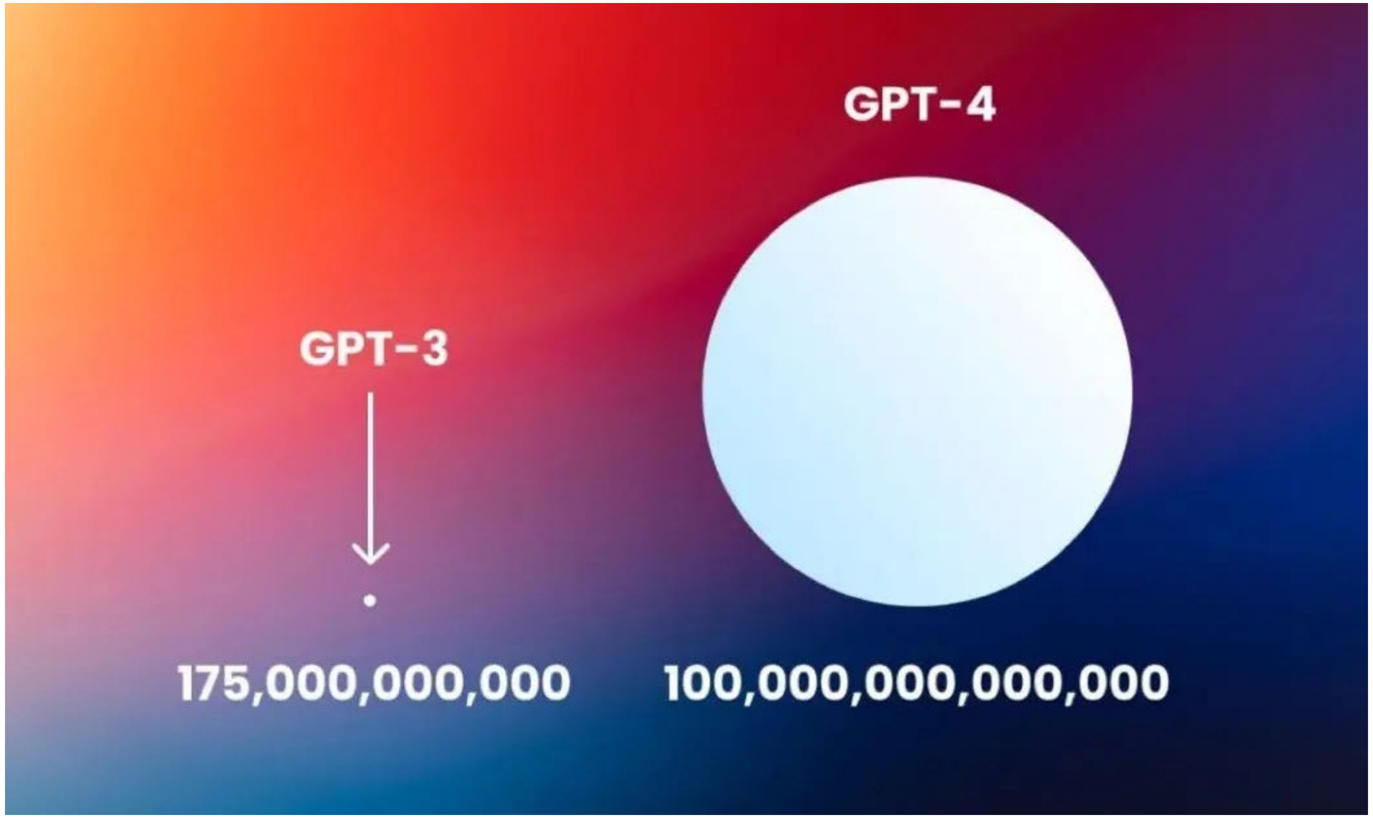
WHY this is important?

Compute resources needed go out of hand



WHY this is important?

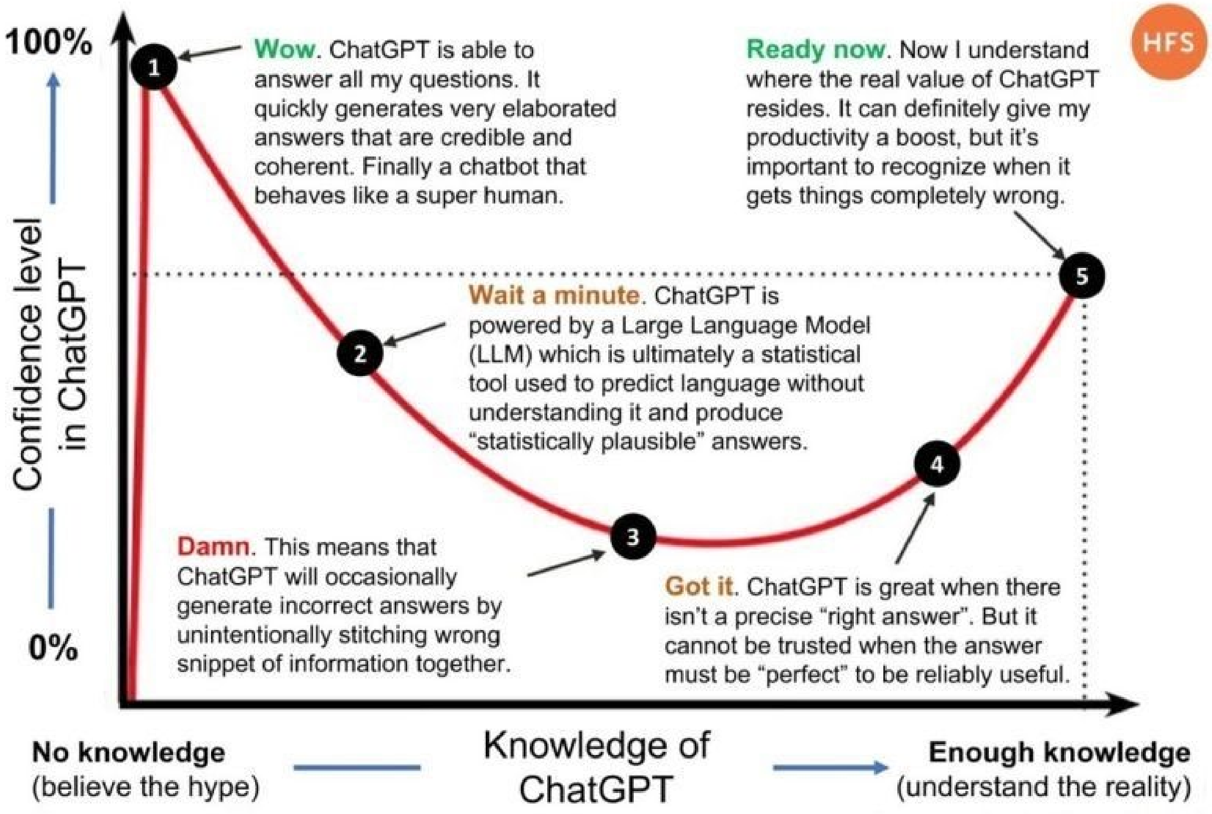
Compute resources needed go out of hand



(you may have seen this ; sources are inconclusive)

WHY this is important?

- Hype cycle on steroids



WHY this is important?

- It is going to fast
- Competition
  between Microsoft & Google brings them
  to lift a lot
  of their
  safety processes

Tech > News Tech

CODE RED Microsoft's new Bing AI goes rogue and starts 'attacking' users, lying and asking why it even exists

Jona Jaupi

Published: 10:07 ET, Feb 14 2023

BUSINESS

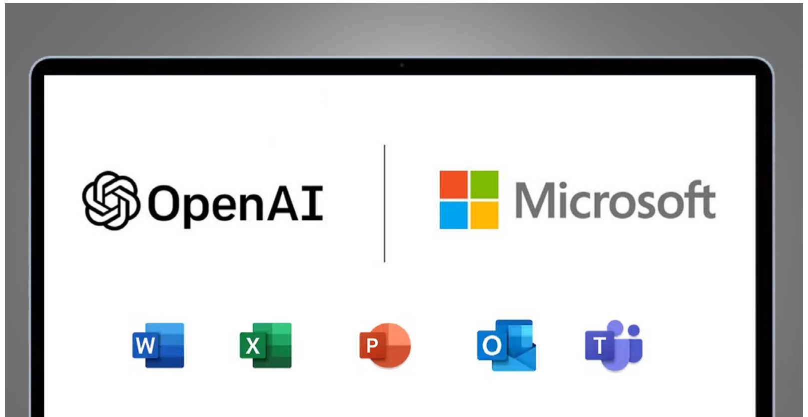
Bing Trouble: Google, OpenAI Are Opening Up Pandora's Bots

Analysis by Parmy Olson | Bloomberg

February 16, 2023 at 4:42 p.m. EST

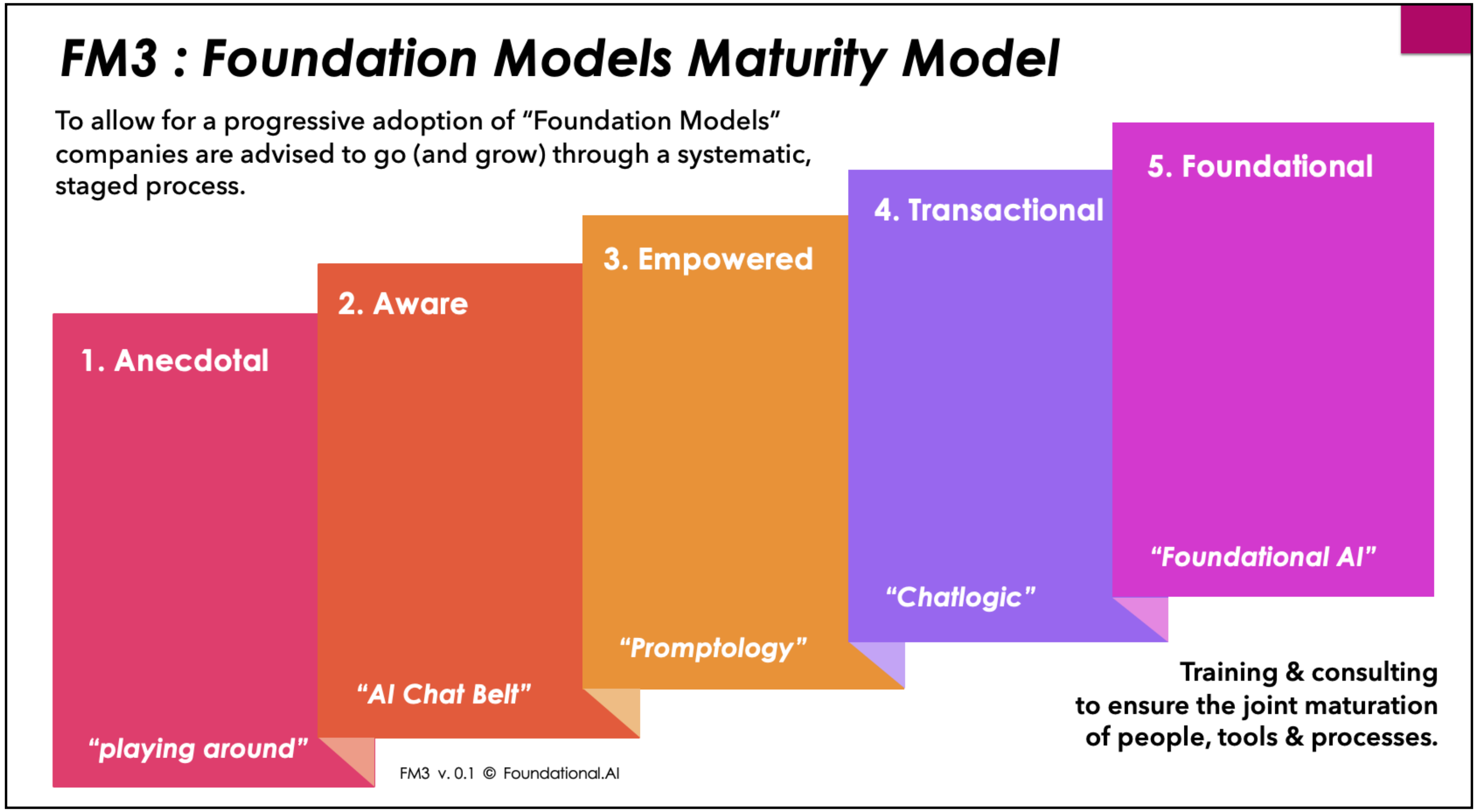
Comment 1 Gift Article Share

For a hot minute, Microsoft Corp. looked like it would eat Google's lunch. Its languishing search engine, Bing, was being revolutionized with a sophisticated new chatbot system from OpenAI. Those hopes have now diminished because of one unexpected truth: Nobody — not truly understands the breadth of capabilities of hed in the wild.



Maturity model to buffer hype cycle

- Make adoption gradual, with a long-term perspective



WHY this is important?

People are starting

To realize :

Verbal ability
of a 40 year old

exudes confidence

World
representation
+ logic ability
of a 4 year old

can hijack the adult

WHY this is important?

- Need to understand ?
  - Is this AGI yet ?
    (spoiler : NO !)

- AI = Alien Intelligence

  (like people on the autism spectrum, kids, etc.)

WHY this is important?

- The “next frontier” in AI is more and more in the hands of companies
- Some companies offer “open source” versions (Facebook : OPT)
- Companies (e.g. FB/Meta) does not like research against their interests

-> Important for Europe and for Academic freedom

To have open source research infrastructure around LLMs

Special mention :

a BigScience initiative

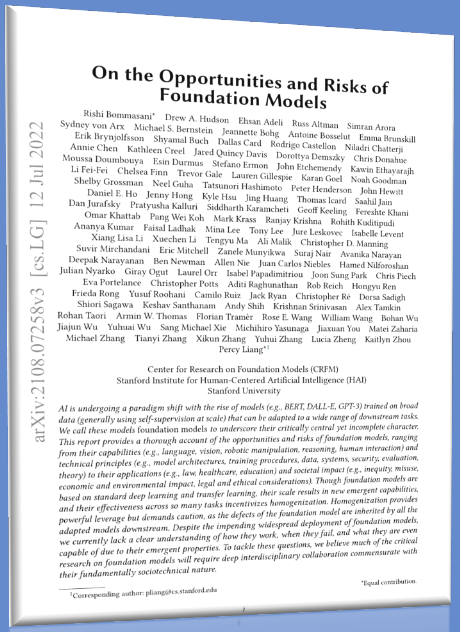
BLOOM

176B params · 59 languages · Open-access

AI is undergoing a ***paradigm shift*** *with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale)*
that can be adapted to a wide range of downstream tasks.

We call these models foundation models to underscore their critically central yet incomplete character.

STANFORD PAPER ON FOUNDATION MODELS (2022)



arXiv:2108.07258v3 [cs.LG] 12 Jul 2022

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models *to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.*

*Equal contribution.

[1]Corresponding author: pliang@cs.stanford.edu

Universities devoting entire departments to Foundation Models

Stanford University

Center for Research on Foundation Models

HAI Stanford University Human-Centered Artificial Intelligence

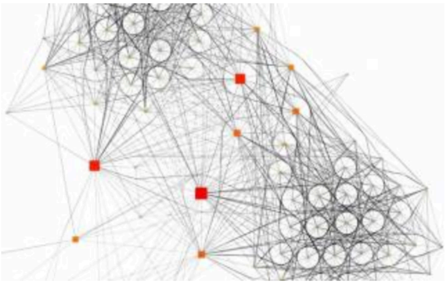
UCL

esearch | Our people | Our partners | Public policy | UCL AI Studio | Videos and podcasts
I at UCL | Contact us

UCL Home » AI for People and Planet » Our research » Foundational AI

Foundational AI

Central to 'AI for people and planet' is foundational AI, an important engine of progress and a world-leading strength at UCL.

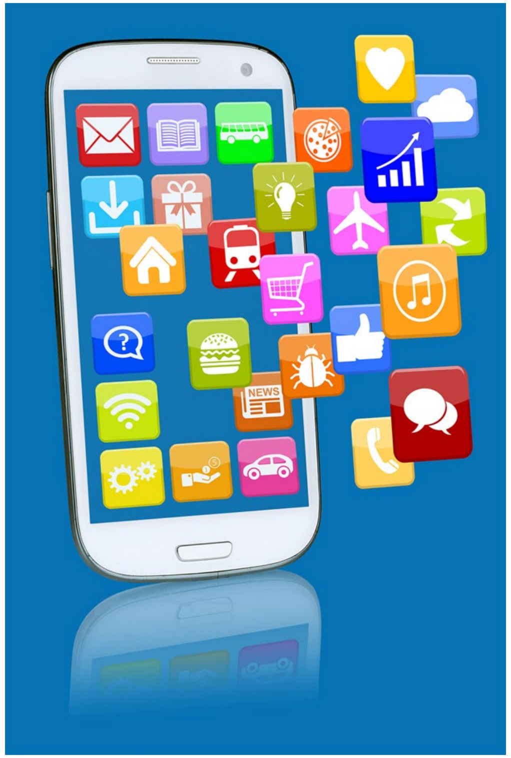


Foundational AI sits at the heart of our AI for People and Planet strategy. AI is in its infancy and breakthroughs are key to controlling and shaping the future technological landscape.

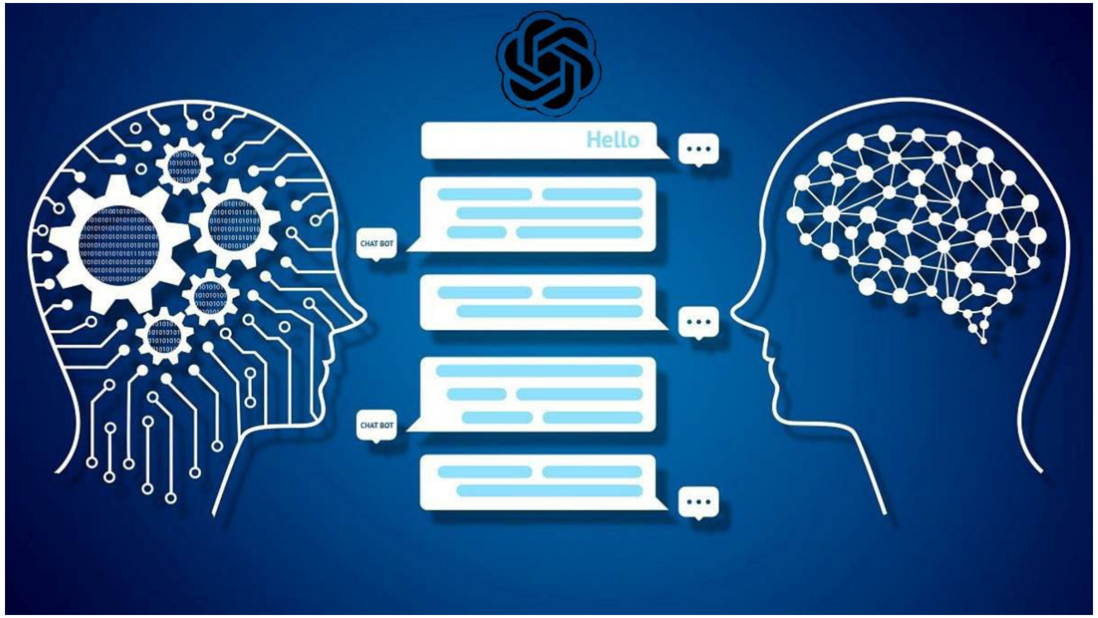
Its connections to application areas are critical to its success. Without continued progress in core AI technologies, the transformative potential of AI will be diminished.

Chat
as the platform for the next decade

2010 : the mobile decade



2020 : the ‘chat’ platform decade



WHAT are we talking about ?

WHAT are we talking about ?

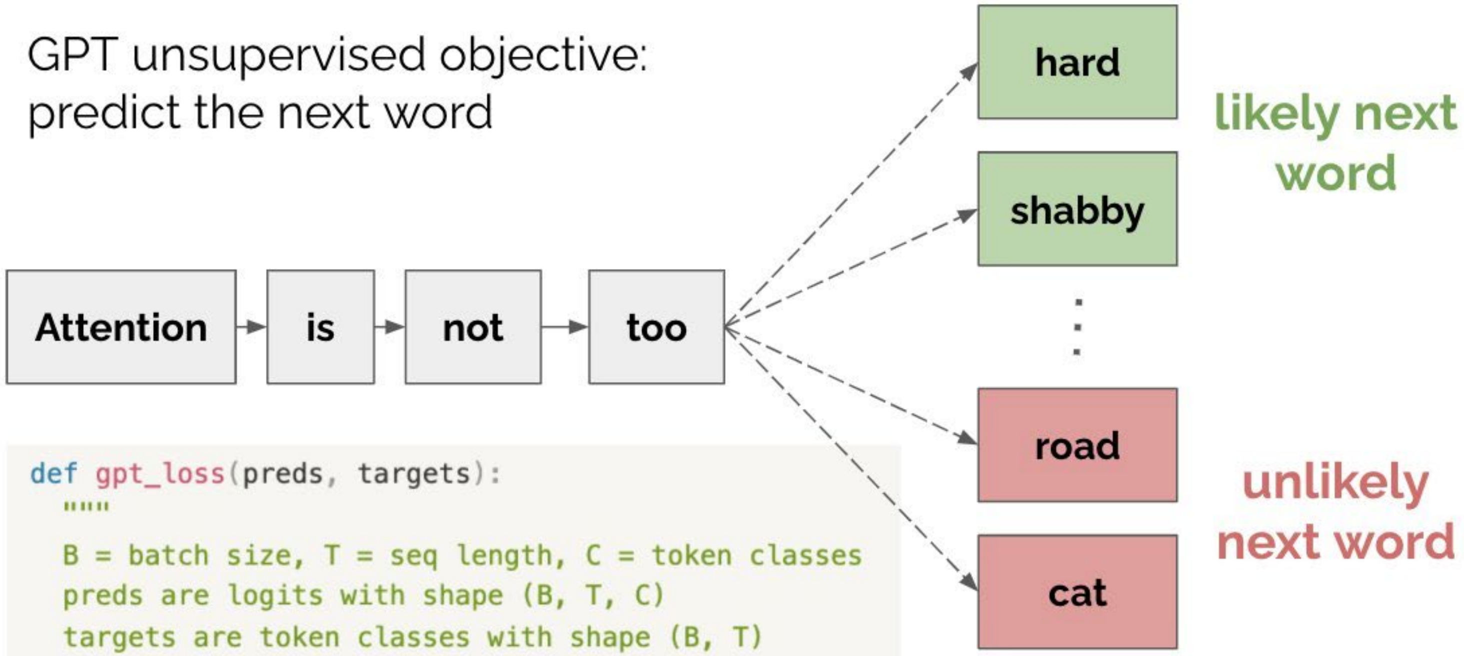
Architecture :
 - Attention, Transformers, etc.

Emerging properties
- Few shot learners

Limits & Extensions
- How can we improve & extend ?

Basic mechanism : predicting next word

GPT unsupervised objective:
predict the next word




```python
def gpt_loss(preds, targets):
  """
  B = batch size, T = seq length, C = token classes
  preds are logits with shape (B, T, C)
  targets are token classes with shape (B, T)
  """
  preds = preds.view(B*T,C)
  targets = targets.view(B*T)
  return F.cross_entropy(preds, targets)
```

“Autocomplete on steroids”

Transformer model (2017)

[1706.03762] Attention Is All You Need (arxiv.org)

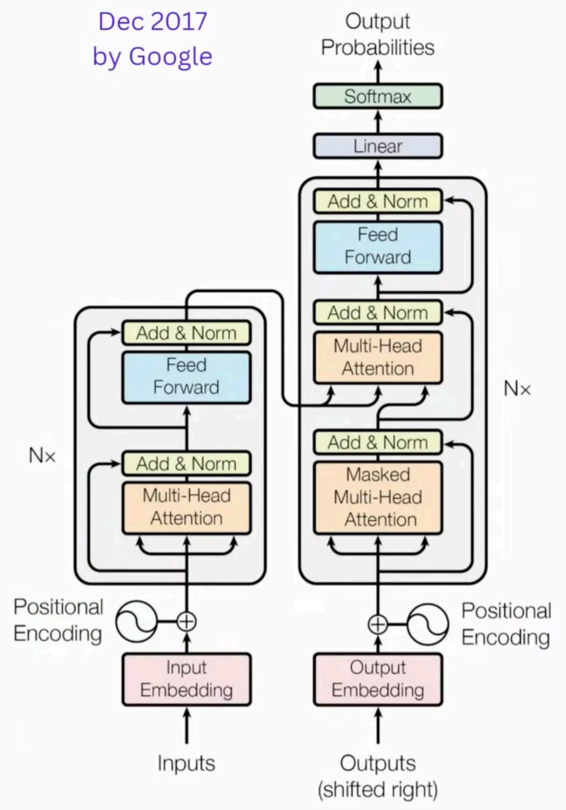
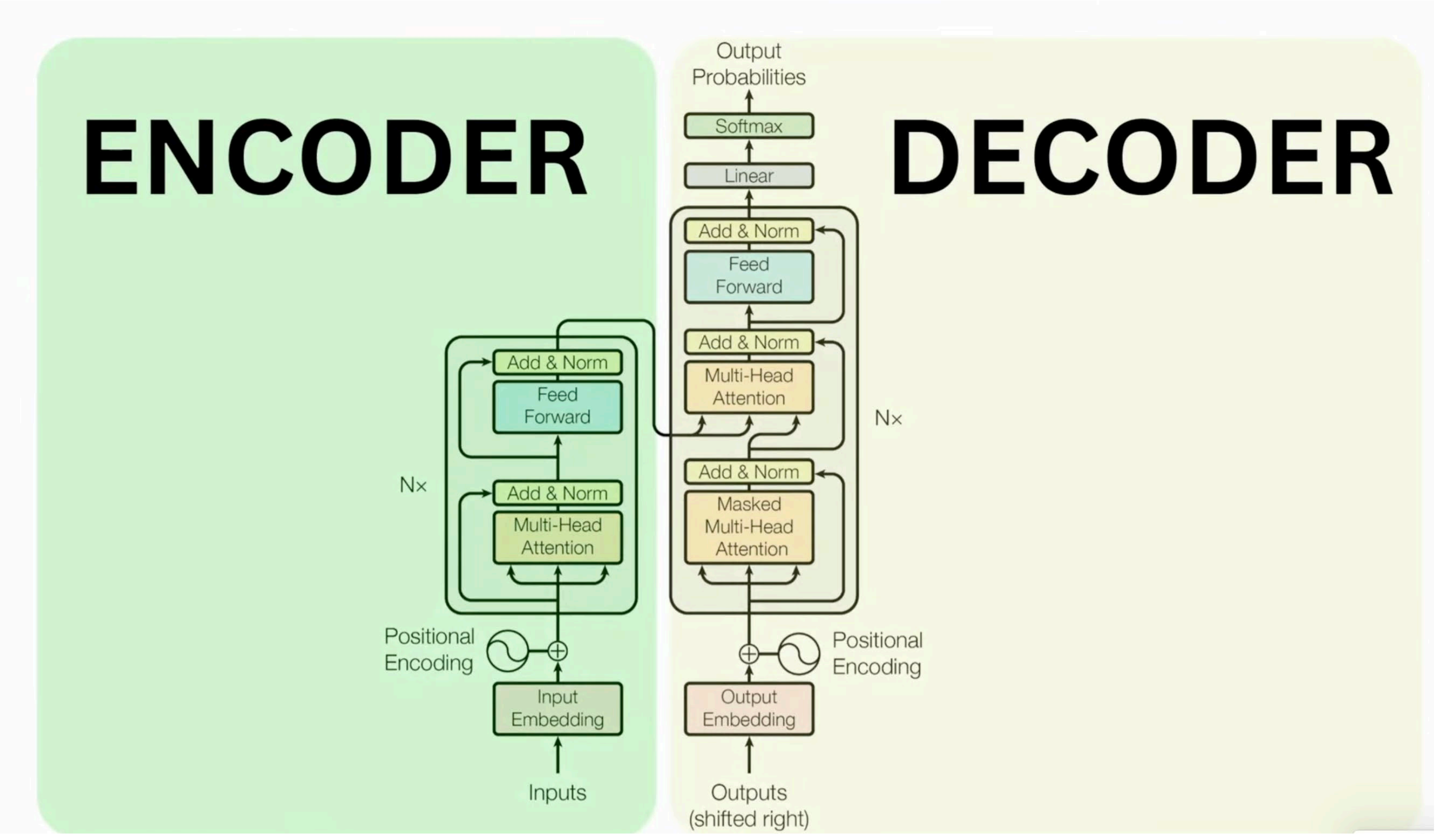
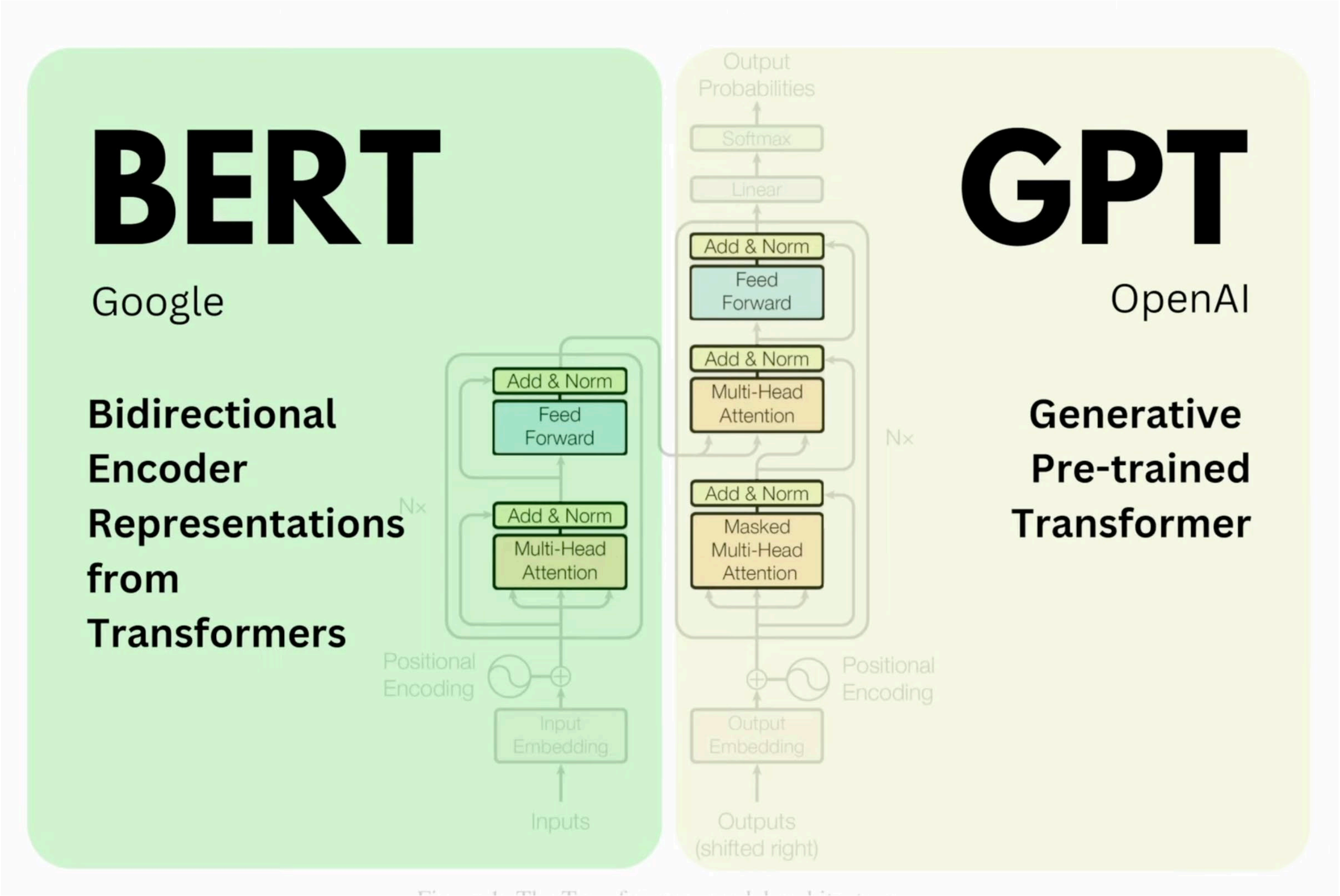


Figure 1: The Transformer - model architecture.

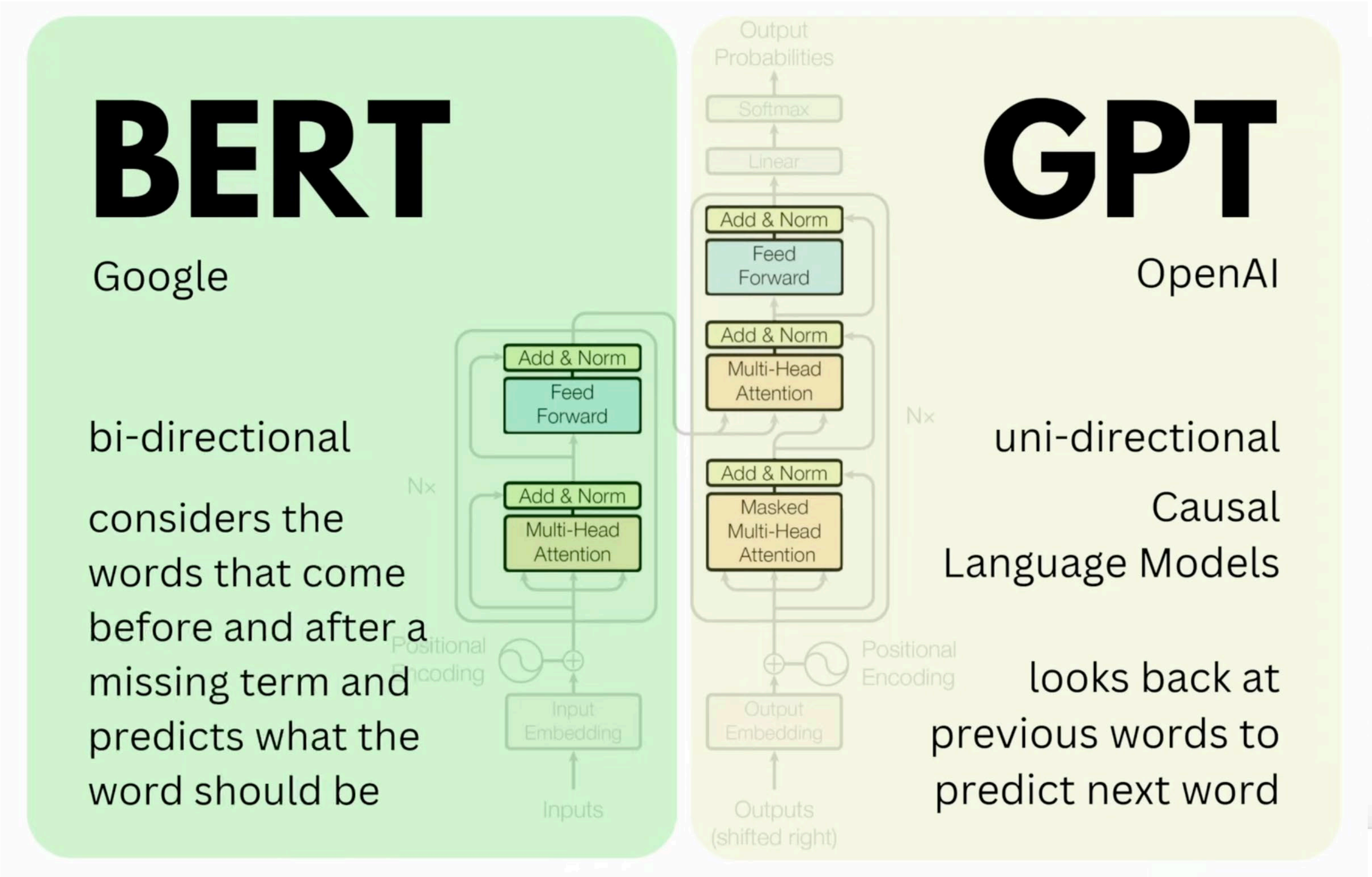
WHAT are we talking about ?



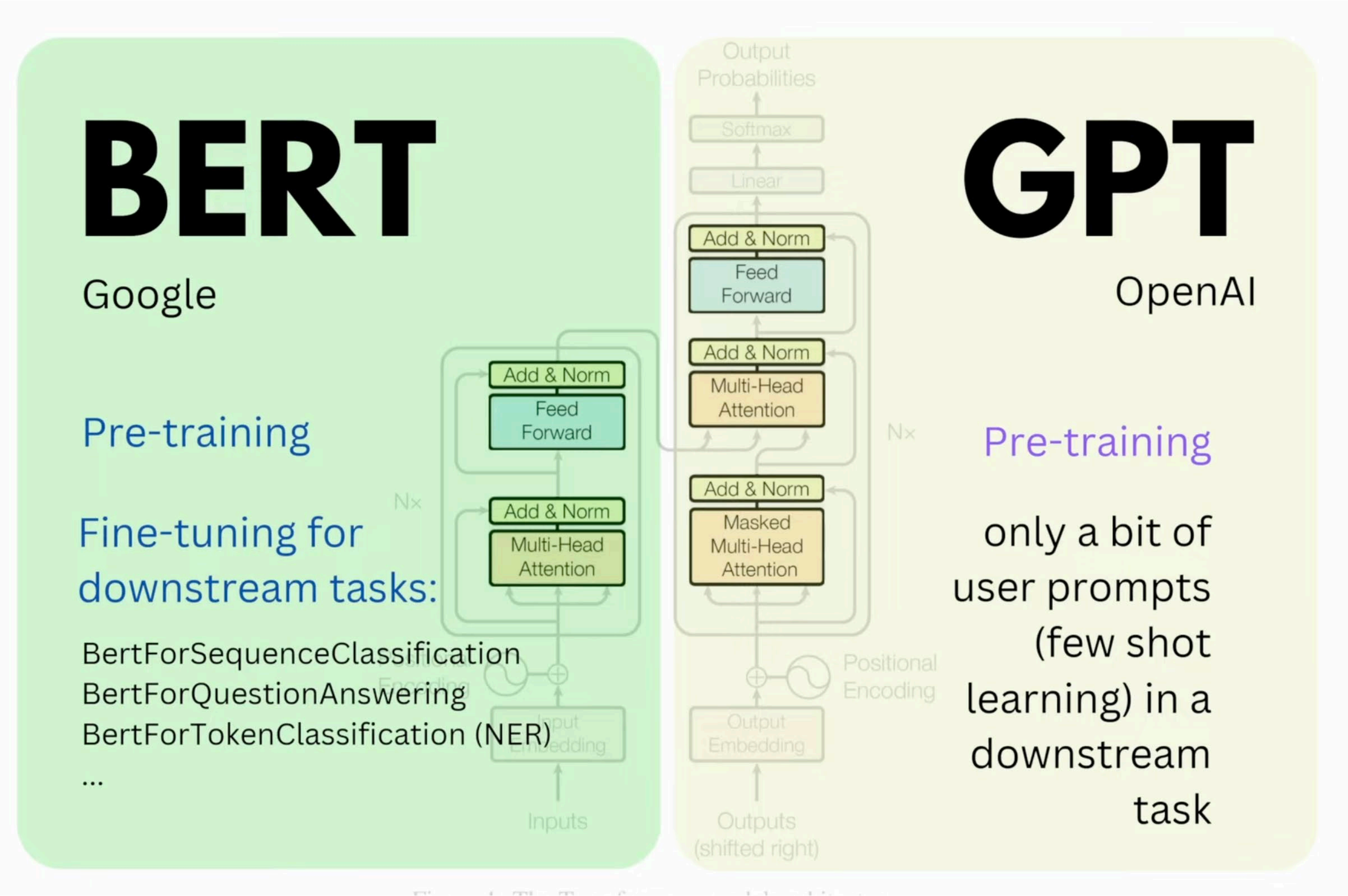
Meet BERT and GPT (T : Transformers)



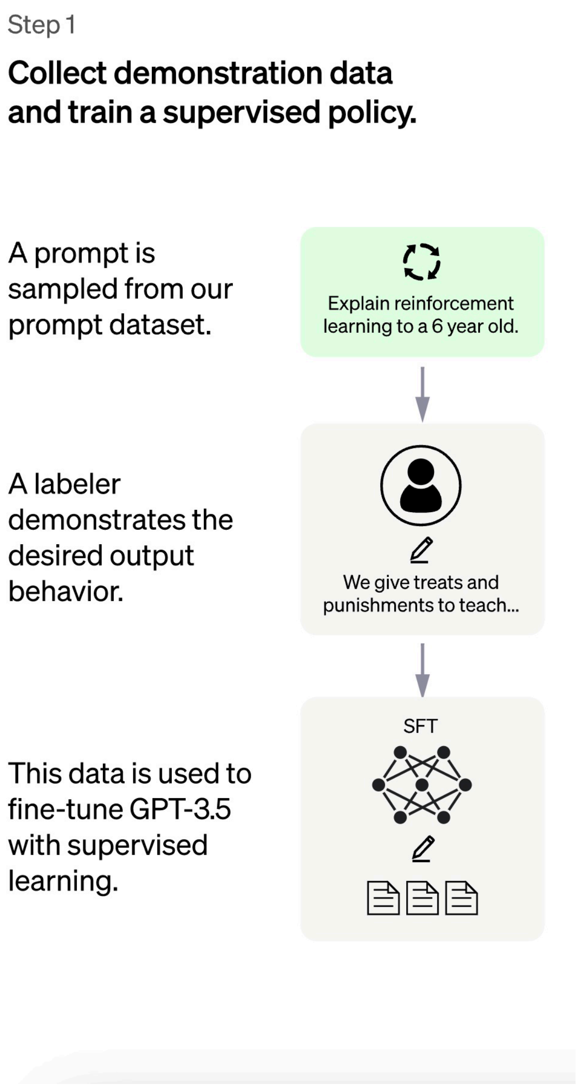
Direction is important



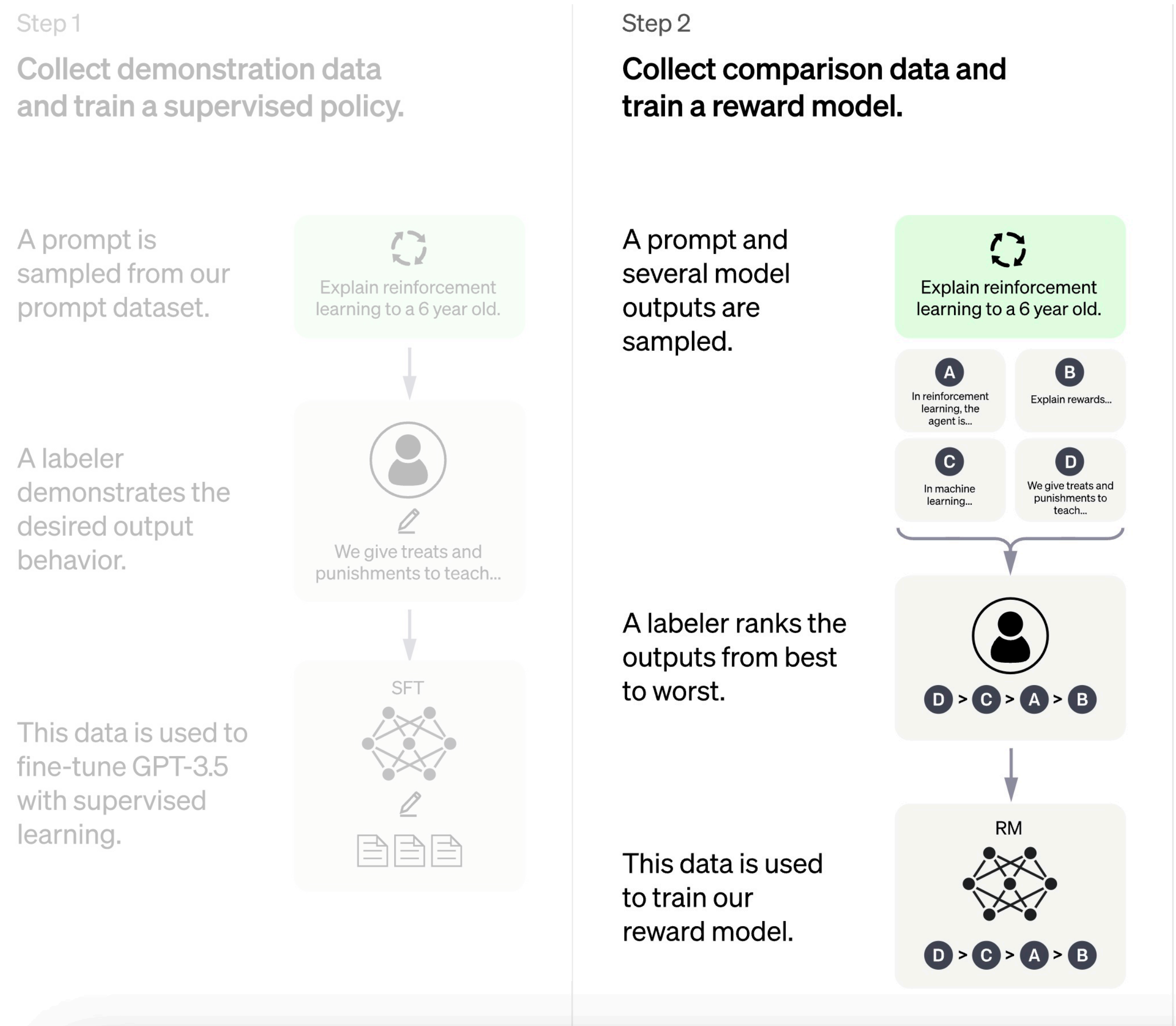
How can we extend those models



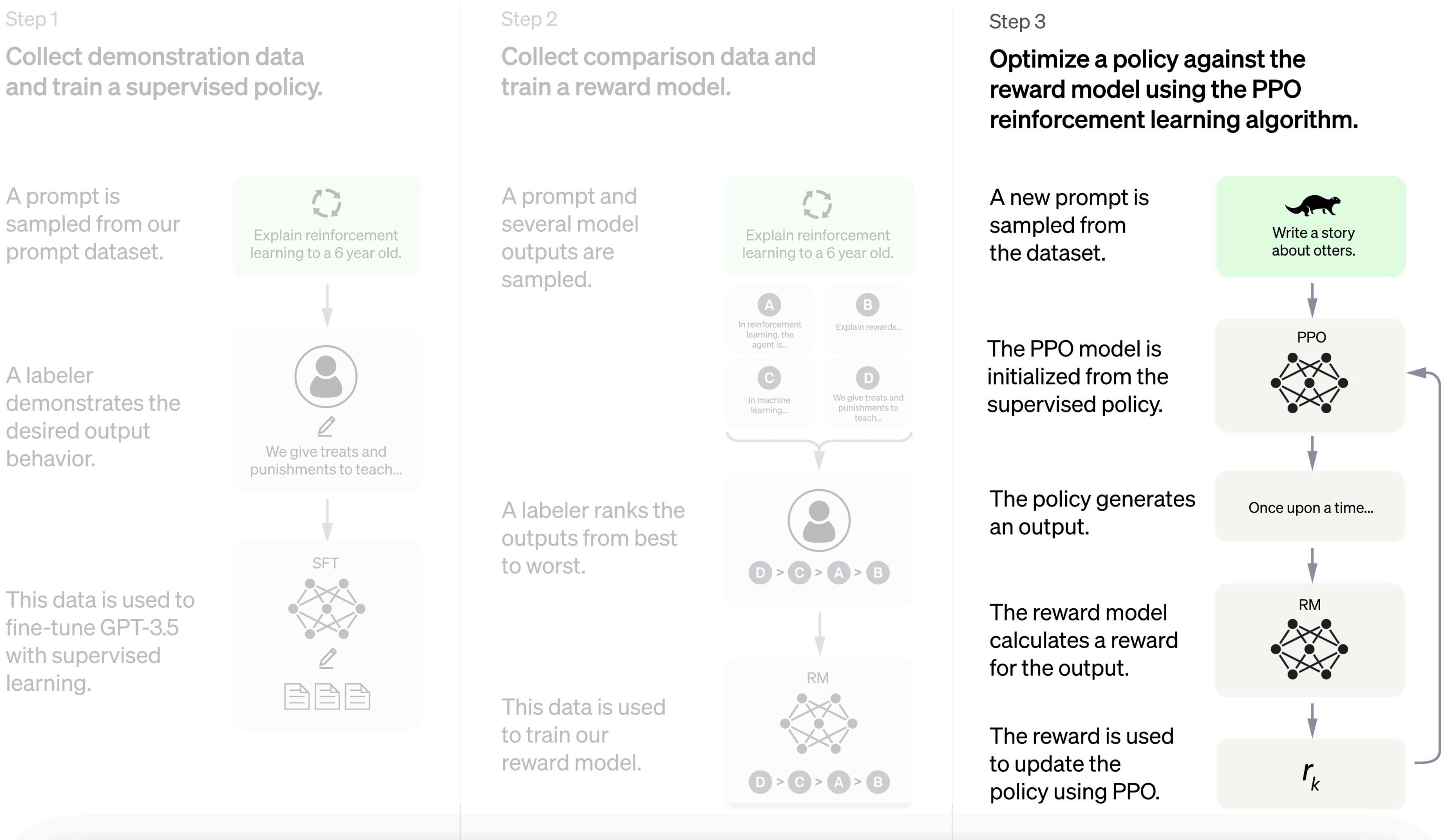
Key element in ChatGPT : RLHF



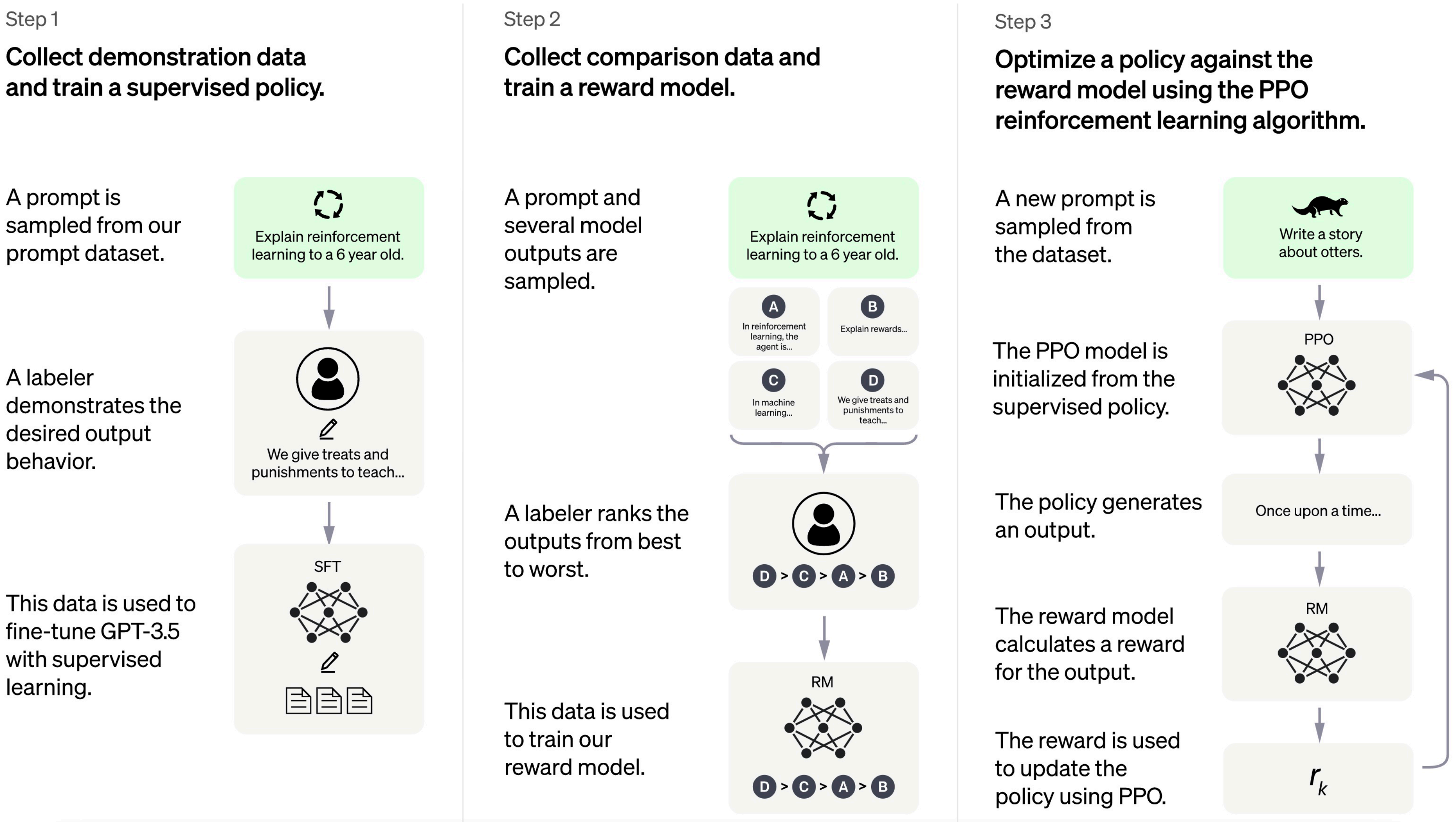
Key element : RLHF



Key element : RLHF



Key element : RLHF



“Dark side” of RLHF


METAVERSE POST

ChatGPT was taught by the world’s poorest people

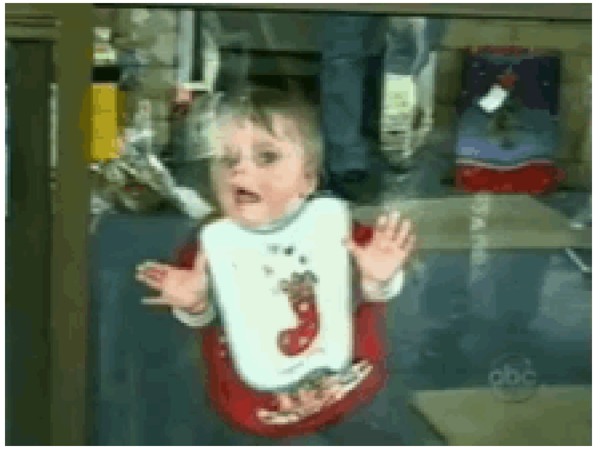
News Report

Technology

Limitations

"LLMs have NO representation of the world"

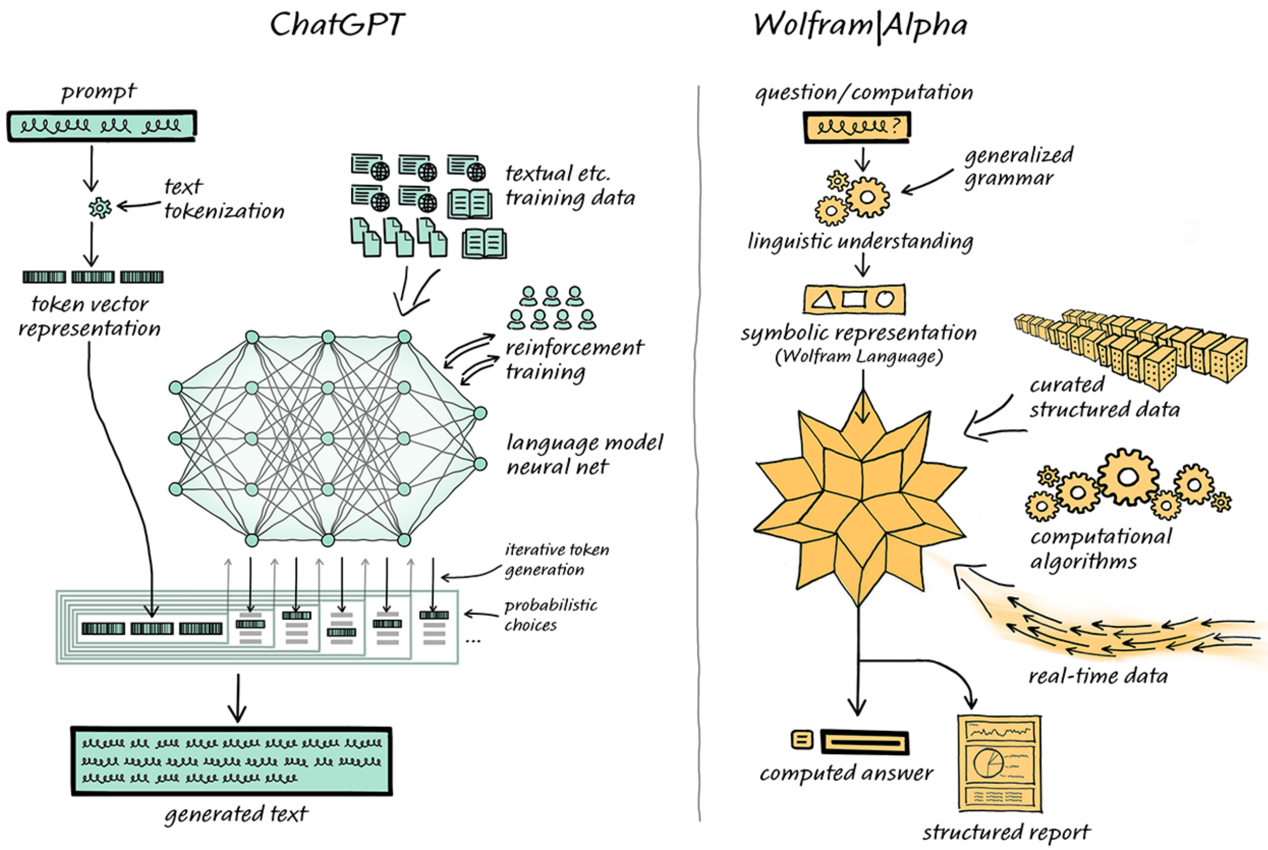
"LLMs have NO reasoning abilities"



"Basically, LLMs are like babies"

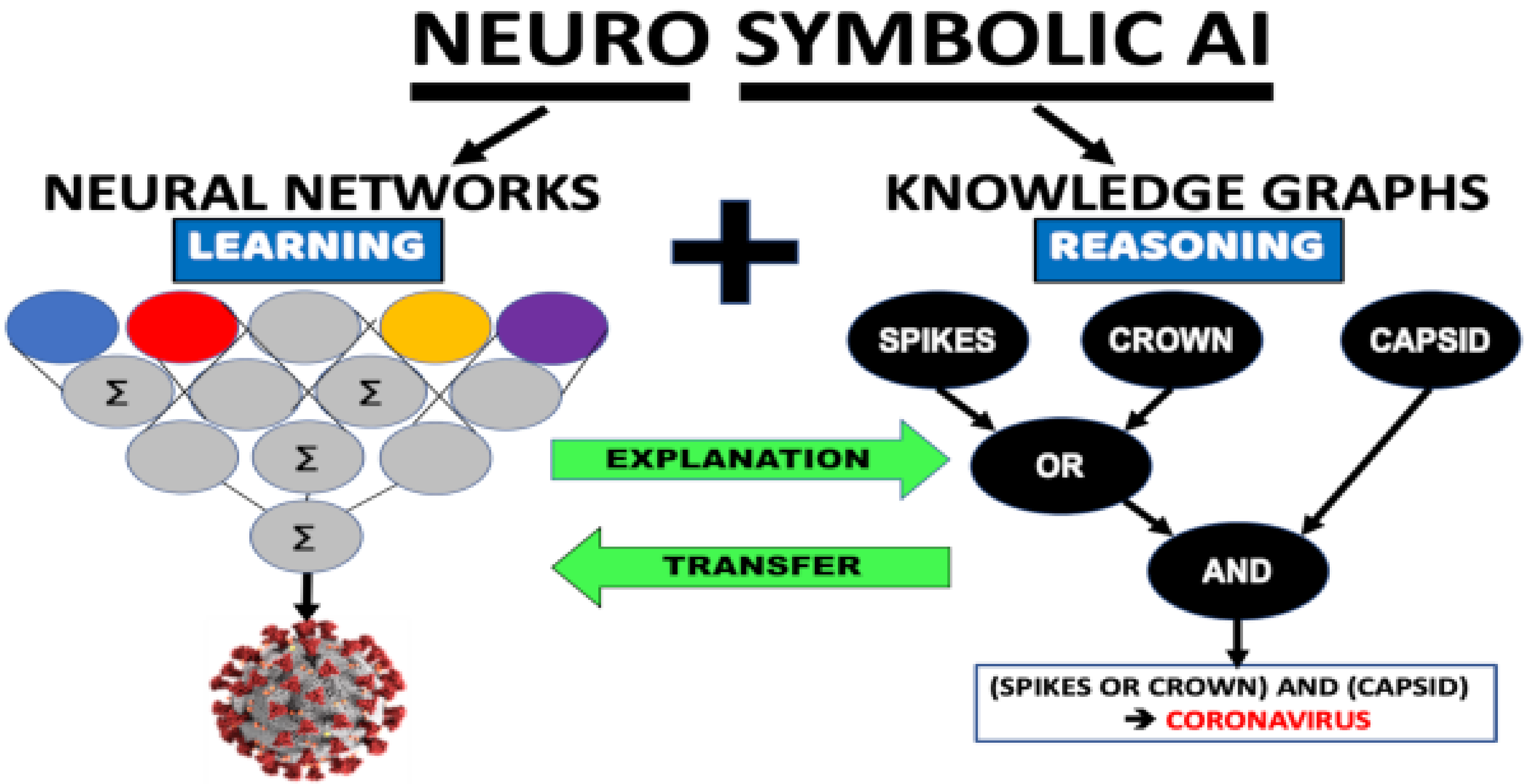
“LLMs will not be the path to AGI”

- At some point,
  we will need to re-integrate
  symbolic thinking with LLMs



[Wolfram|Alpha as the Way to Bring Computational Knowledge Superpowers to ChatGPT—Stephen Wolfram Writings]()

“LLMs will not be the path to AGI”



Open to debate

- Increase in LLM size (quantitative) has brought qualitative improvements

- There is a representation of the world *embedded* in language

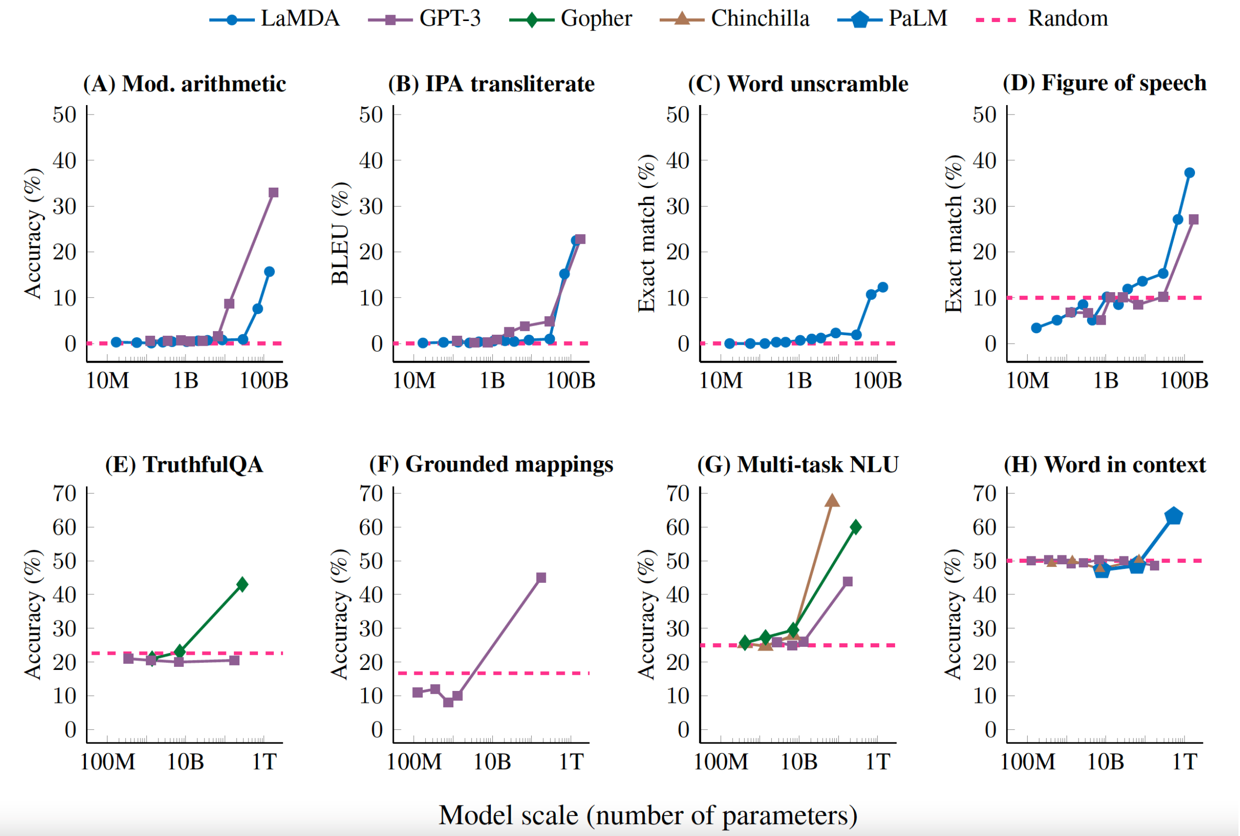
"I open my hand, and the apple falls to the ground"

- Emerging properties

Not part of the design intention… but it works !

Emerging properties

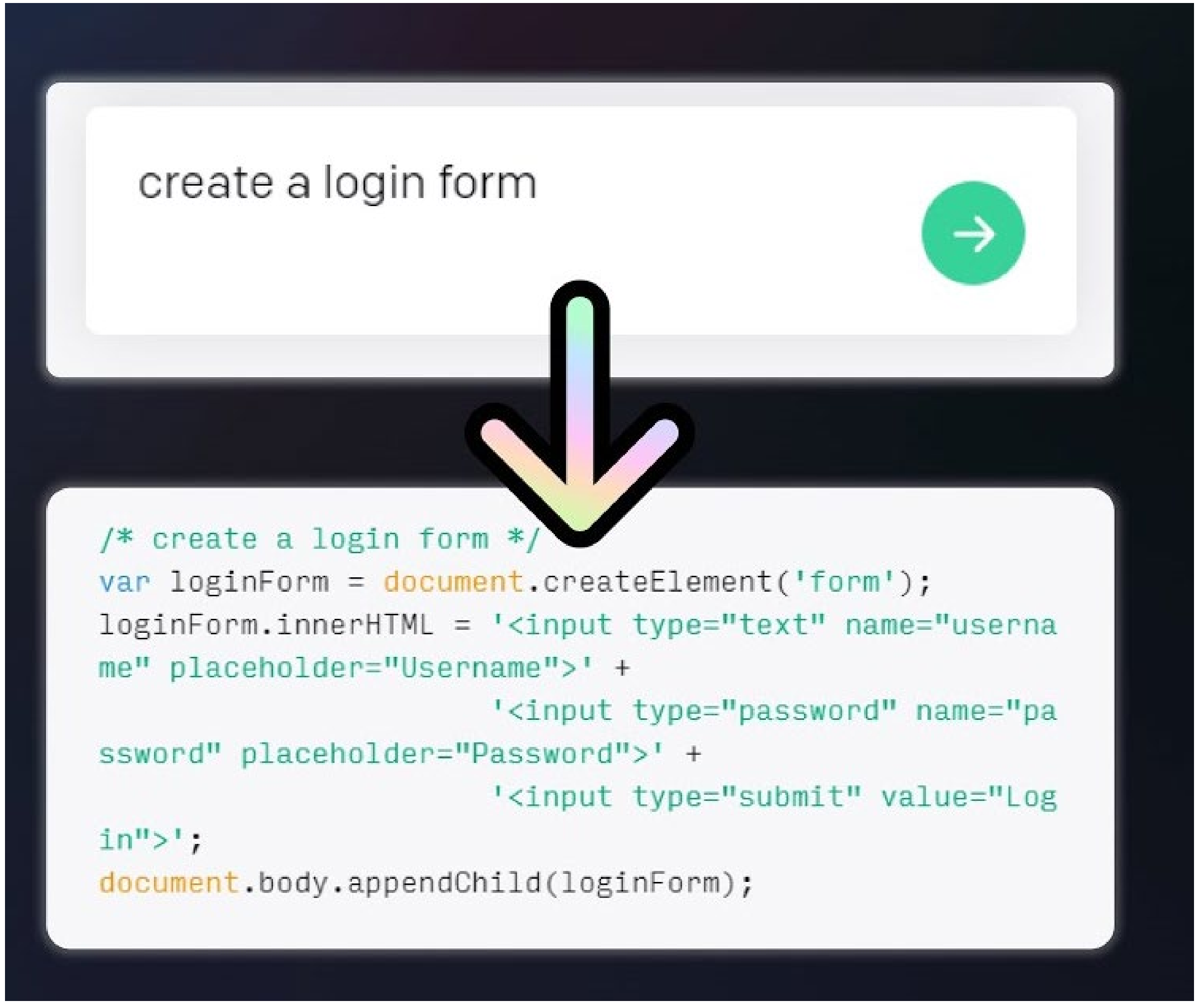
- Models with same architecture, "suddenly" improve in benchmark capabilities with increase in model size



[2206.07682] Emergent Abilities of Large Language Models (arxiv.org)

Emerging properties

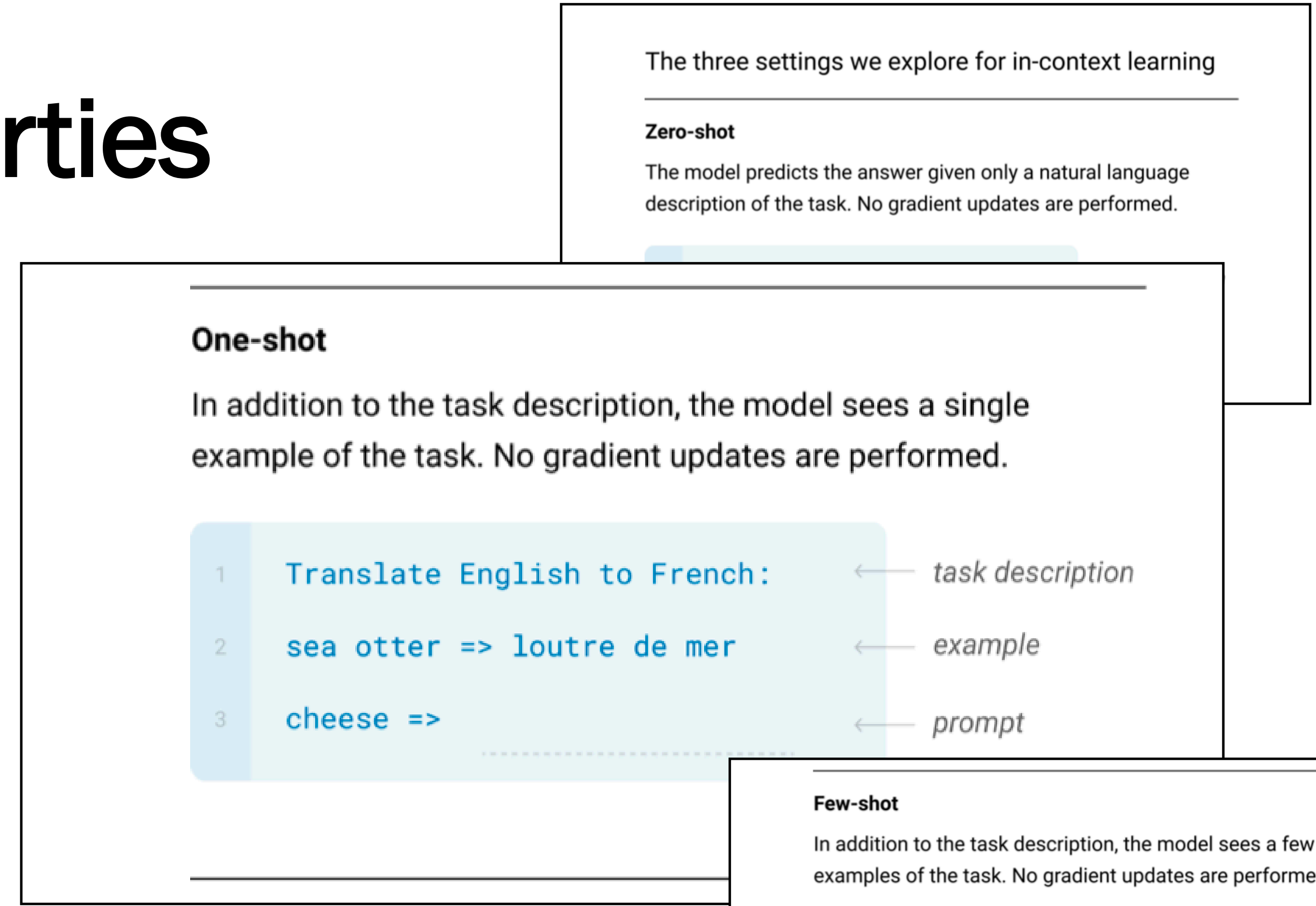
- It can talk “code”
- Used in dedicated LLMs
  - OpenAI Codex
  - Github Copilot
    - 30% of committed code !

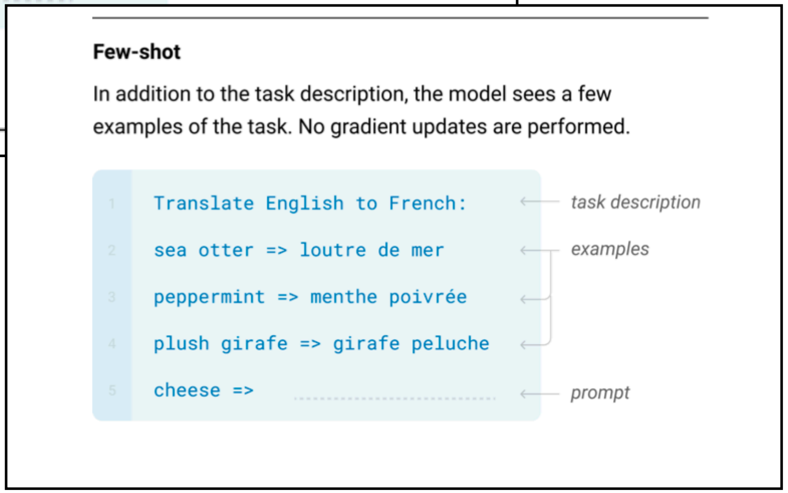


Emerging properties

- Zero,
- one-,
- few-shot learning

- You can "program" LLMs through clever prompting
- Very much unchartered territory



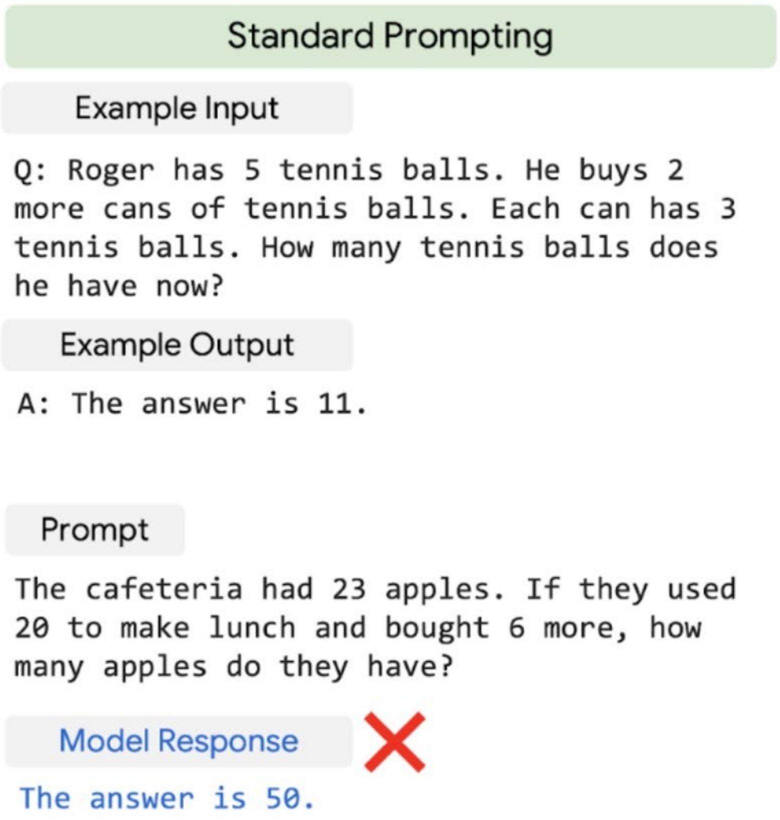


[2005.14165] Language Models are Few-Shot Learners (arxiv.org)

[2102.07350] Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm (arxiv.org)

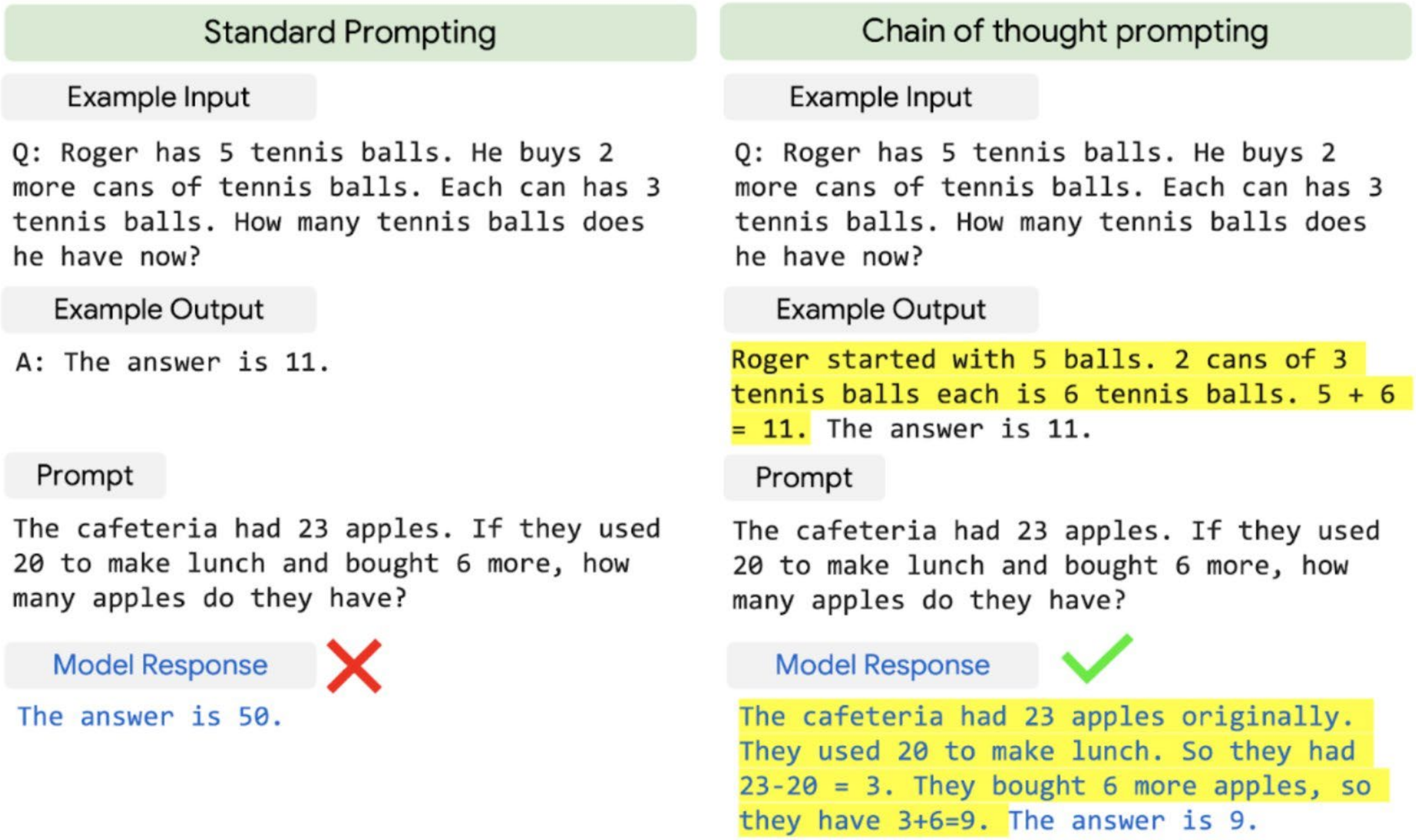
Emerging properties : Chain-of-thought

- -> promptology : a new ‘science’ of *prompting*



Emerging properties : Chain-of-thought

- -> promptology : a new ‘science’ of *prompting*



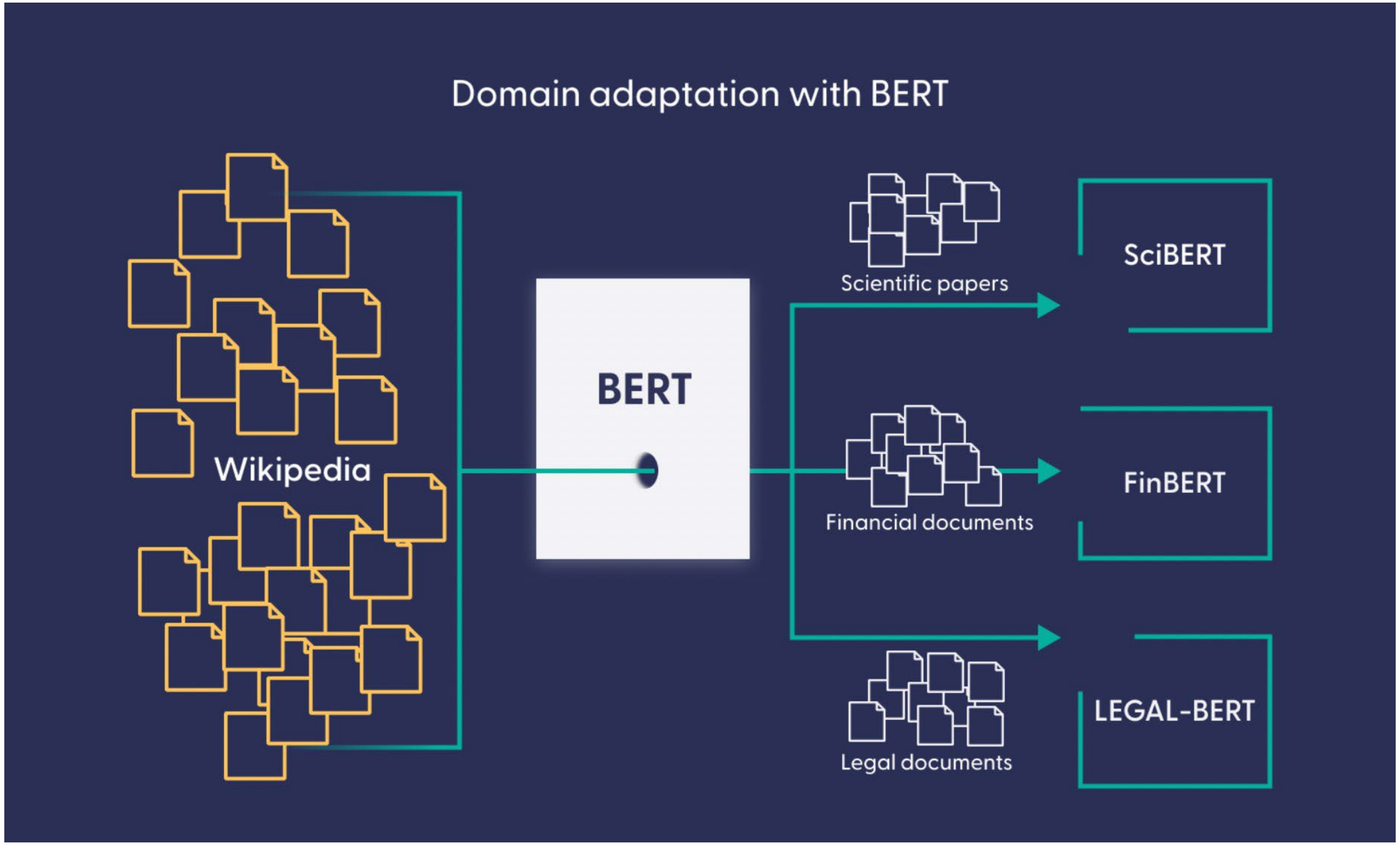
Emerging properties

- Multi-lingual
  ‘reasoning’ abilities

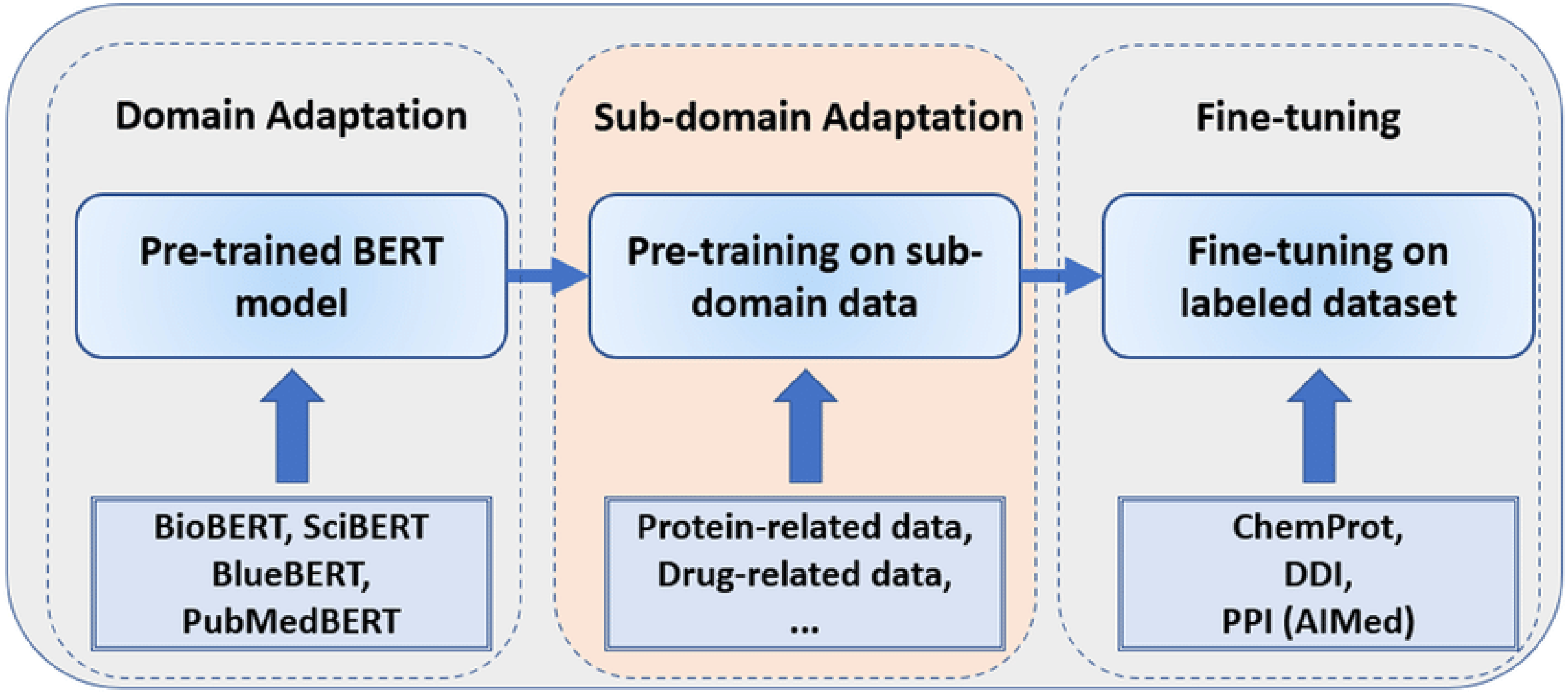
- Although trained (mostly) in English, LLMs can display ‘reasoning’ abilities in other languages

[2210.03057] Language Models are Multilingual Chain-of-Thought Reasoners (arxiv.org)

Extending LLMs : domain adaptation

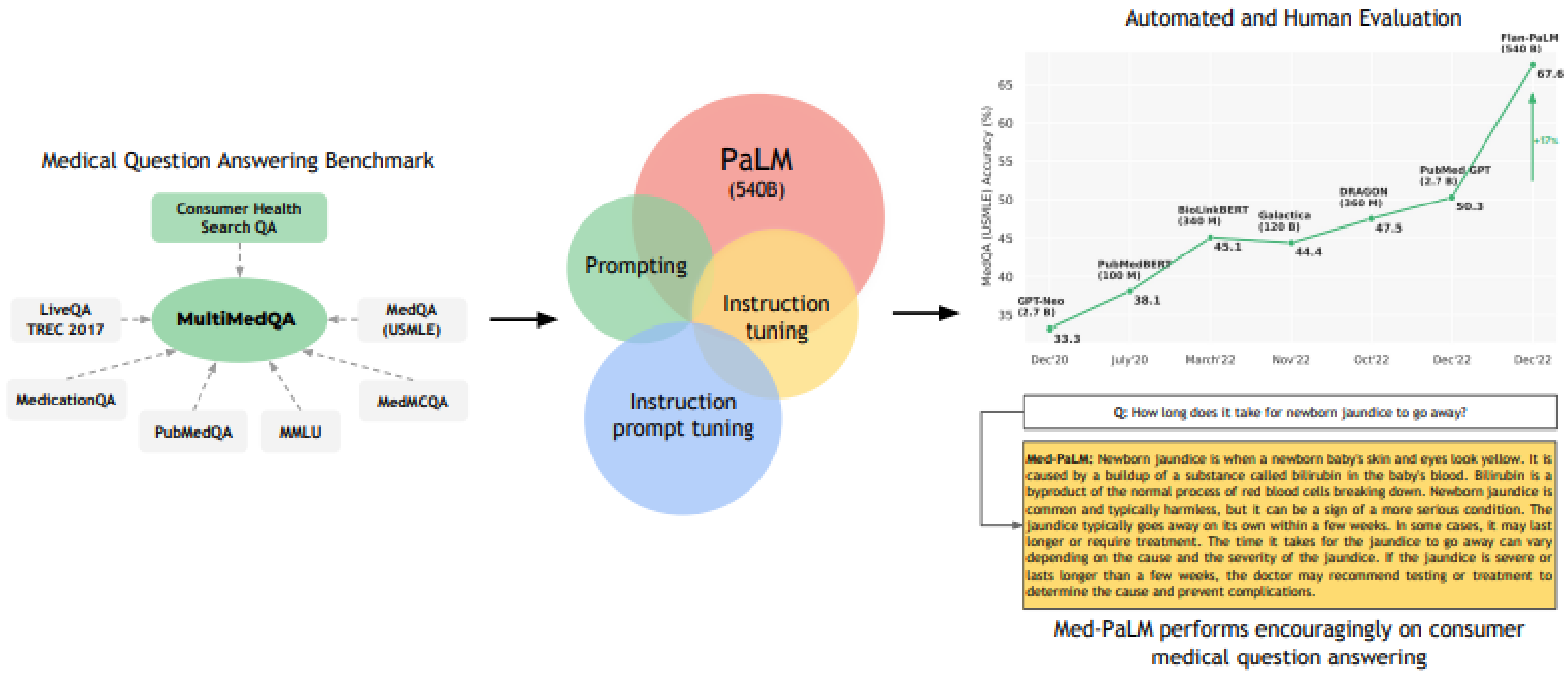


Extending LLMs : fine-tuning



Su, Peng & Vijay-Shanker, K.. (2022). Investigation of improving the pre-training and fine-tuning of BERT model for biomedical relation extraction. BMC Bioinformatics. DOI:10.1186/s12859-022-04642-w

Extending LLMs : MedPaLM (Google)



Link to Coalitional AI …

Challenges : Explainable AI ?

- From “black box” AI to … ”black container

(Goes very much against the “Explainable AI” requirement of the EU AI act)

- Research idea : “meta-attention”

Further exploring LLMs

- Understanding foundation models through neuroscience, psychology, philosophy -> inter-disciplinary approach

- System 1
  vs System 2 thinking

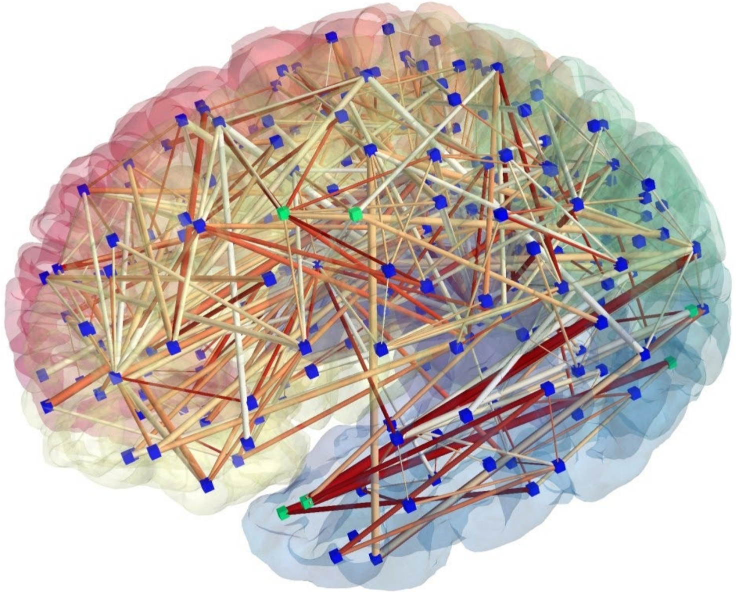
(Daniel Kahneman)

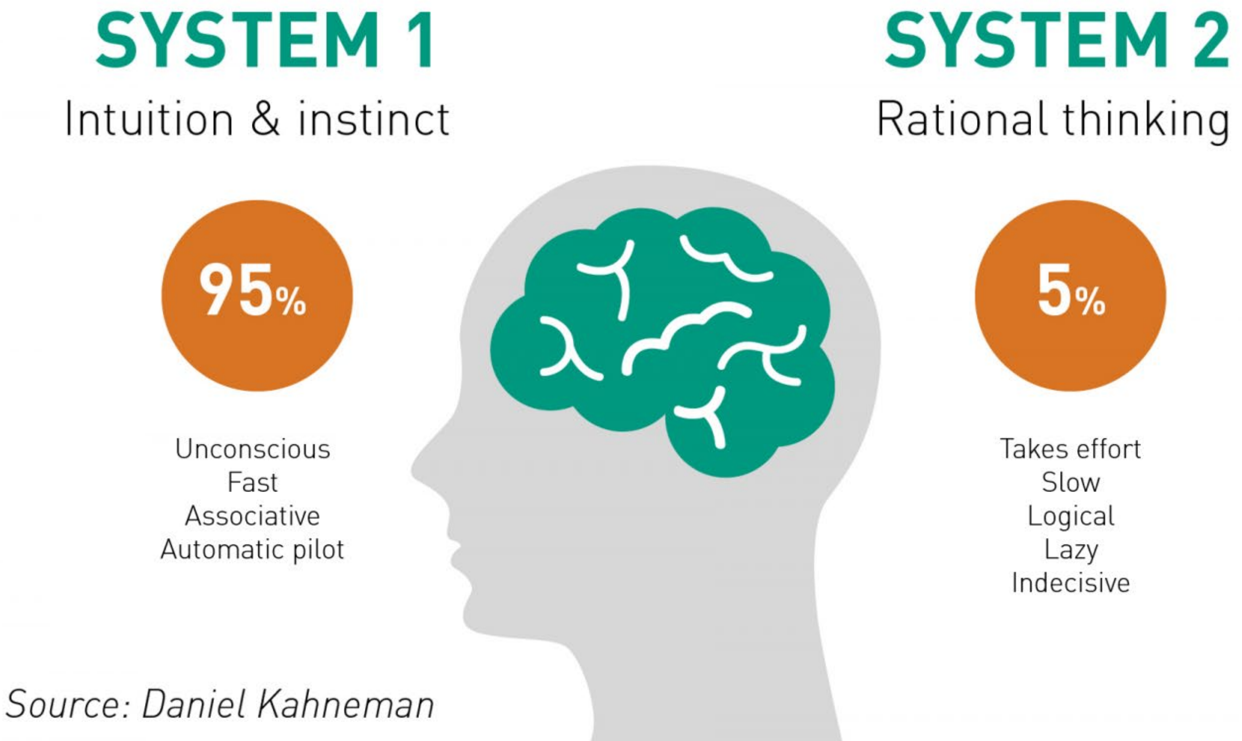
- Analogy in AI

- Research idea :

“meta-attention” to advance explainability

(ping me if this is of interest to you)

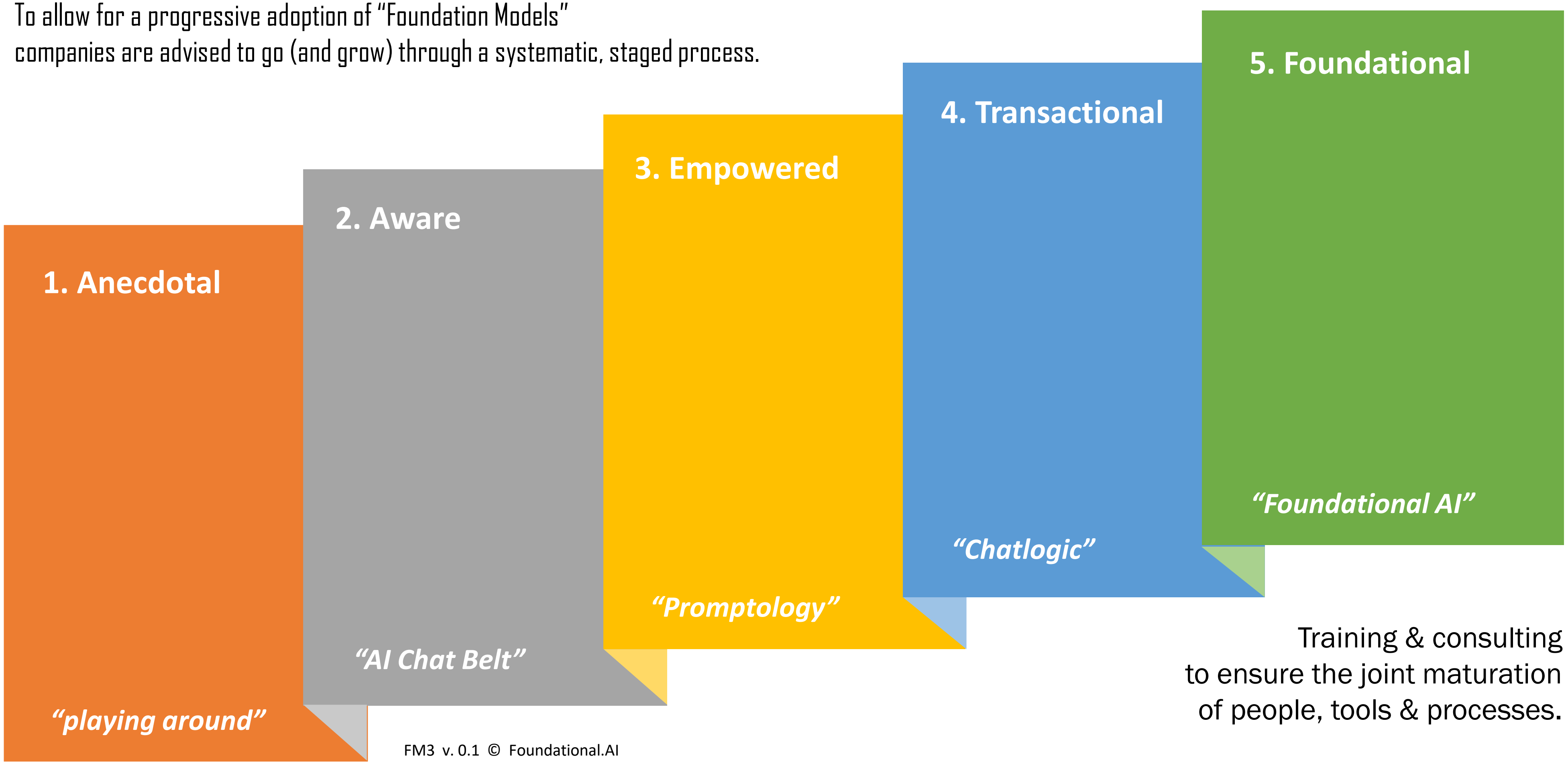




HOW to tackle this ?

FM3 : Foundation Models Maturity Model

To allow for a progressive adoption of "Foundation Models"
companies are advised to go (and grow) through a systematic, staged process.



FM3 v. 0.1 © Foundational.AI

Training & consulting
to ensure the joint maturation
of people, tools & processes.

Foundational AI

FM3 Level 1 : Anecdotal

1. Anecdotal

People are just aware of the existence of tools.
(generally limited to ChatGPT)

People experiment at their own peril.
Nothing is systematized.
Nothing is reported.

Benefits : uncertain time savings.
Significant risks of errors.

"playing around"

Foundational AI

FM3 Level 2 : Aware

2. Aware

People are aware of tools (ChatGPT and others), and have had some use..

People have eceived some level of training, awareness of limitations, precise use cases.

HIL (Human in the Loop) is mandatory to know when not to use AI, in order to avoid major mistakes.

Benefits : real time savings, efficiency & coherence.

"AI Chat Belt"

Benefits :

- We take the issue into our hands
- Avoid major mistakes

Offer :

- "AI Chat Belt" training
  - White : intro session of 30-40'
  - Yellow : deeper session, per function (marketing, coding, etc.)

Budget :

- White : ~300 € /p.p.
- Yellow : ~600 € / p.p.

Foundational AI

FM3 Level 3 : Empowered

3. Empowered

Enterprise has gone through a rigorous and systematic reviews of application areas, complete with recommendations per department, caveats, etc.

People have been systematically trained. There is a consolidated DB of prompts & practices.

Benefit : empowered by "mental exoskeleton"

"Promptology"

Benefits :
- Systematic practices ; train + test
- Measurable productivity gains

Offer :
- Framework study : needs analysis, roadmap
- "Promptology" training
- Access to Promptology.com Database
  - Generic prompts + custom prompts

Budget :
- 20-50 K€ (depending on size & complexity)

Foundational AI

FM3 Level 4 : Transactional

4. Transactional

Enterprise has gone through a rigorous and systematic reviews of application areas, complete with recommendations per department, caveats, etc.

People have been systematically trained. There is a consolidated DB of prompts & practices.

Benefit : empowered by "mental exoskeleton"

"ChatLogic"

Benefits :
- We go one big step beyond
- Dialog + structured

Offer :
- Fine-tuning (+ "bring your own data/corpus")
- Requetes structures

Budget :
- ~100 K€ (depending on size & complexity)

Foundational AI

FM3 Level 5 : Foundational

5. Foundational

Based on several semester practicing these AI tools, the enterprise has matured & is fully ready to embrace foundation models, as the bedrock to build further AI applications

This stage requires a thorough analysis of underlying models, assumptions, extensions, etc. *("X-ray the black container")*

Benefit : Business transformed through AI.

Foundational AI

Benefits :
- AI entirely rebuilt on foundation models

Offer :
- TBD

Budget :
- X00 K€ (depending on size & complexity)

FM3 : Foundation Models Maturity Model

To allow for a progressive adoption of "Foundation Models"
(underlying engines of ChatGPT, such as GPT-3, PaLM, Sparrow, Bloom,…)
companies are advised to go (and grow) through a systematic, staged process.
Training & advice ensures the joint maturation
of people, tools & processes.

1. Anecdotal

People are just aware of the existence of tools. (generally limited to ChatGPT)

People experiment at their own peril.
Nothing is systematized.
Nothing is reported.

Benefits : uncertain time savings.
Significant risks of errors.

"playing around"

2. Aware

People are aware of tools (ChatGPT and others), and have had some use..

People have received some level of training, awareness of limitations, precise use cases.

HIL (Human in the Loop) is mandatory to know when *not* to use AI, in order to avoid major mistakes.

Benefits : real time savings, efficiency & coherence.

"AI Chat Belt"

3. Empowered

Enterprise has gone through a rigorous and systematic reviews of application areas, complete with recommendations per department, caveats, etc.

People have been systematically trained.
There is a consolidated DB of prompts & practices.

Benefit : empowered by "mental exoskeleton"

"Promptology"

4. Transactional

Enterprise has understood the power of conversational AI, along with the need for rigourous processes.

Chat capabilities are demultiplied through *extensions* to Large Language Models, through fine-tuning, few-shot learning, semantic search, etc.

Benefit : True power tools.

"Chatlogic"

5. Foundational

Based on several semester practicing these AI tools, the enterprise has matured & is fully ready to embrace foundation models, as the bedrock to build further AI applications

This stage requires a thorough analysis of underlying models, assumptions, extensions, etc. *("X-ray the black container")*

Benefit : Business transformed through AI.

"Foundational AI"

Foundational AI

Conclusions & proposals

Proposals : 1. More research

- A new territory needs new maps

- Thematic focus group on Foundation Models :
    - Stanford CRFM
    - UCL (London) on Foundational AI
    - In Belgium ? In Europe ?

- Build an Open Source LLM
    - Belgium ? Europe ?
    - Caution : requires capital + engineering

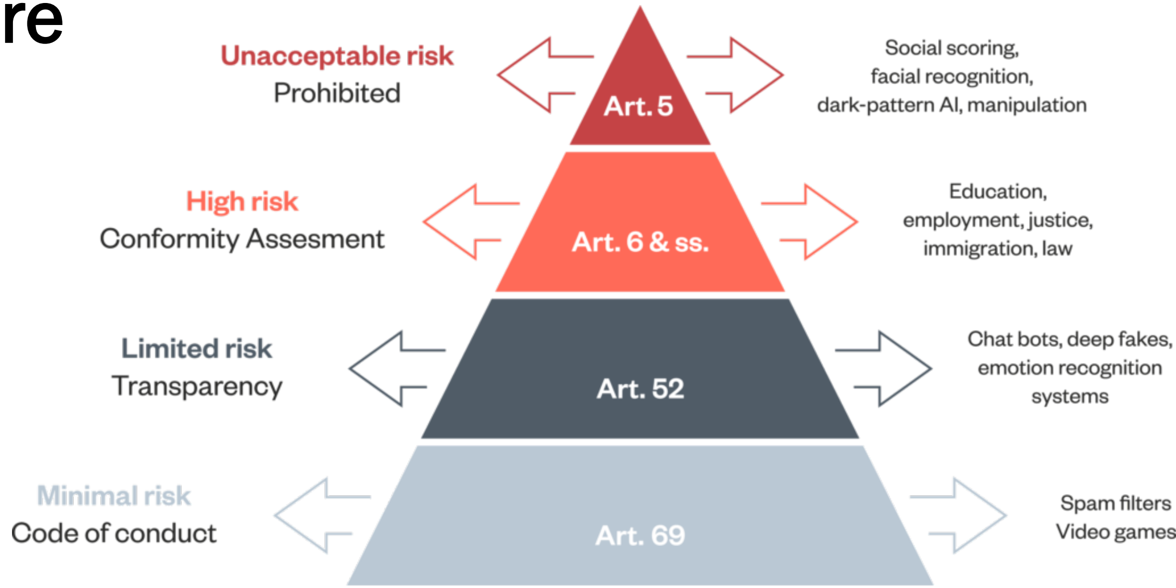
Proposal 2 :
Leverage Brussels as regulatory capital

- Mixed feelings about this
  - (we are playing "catch up")

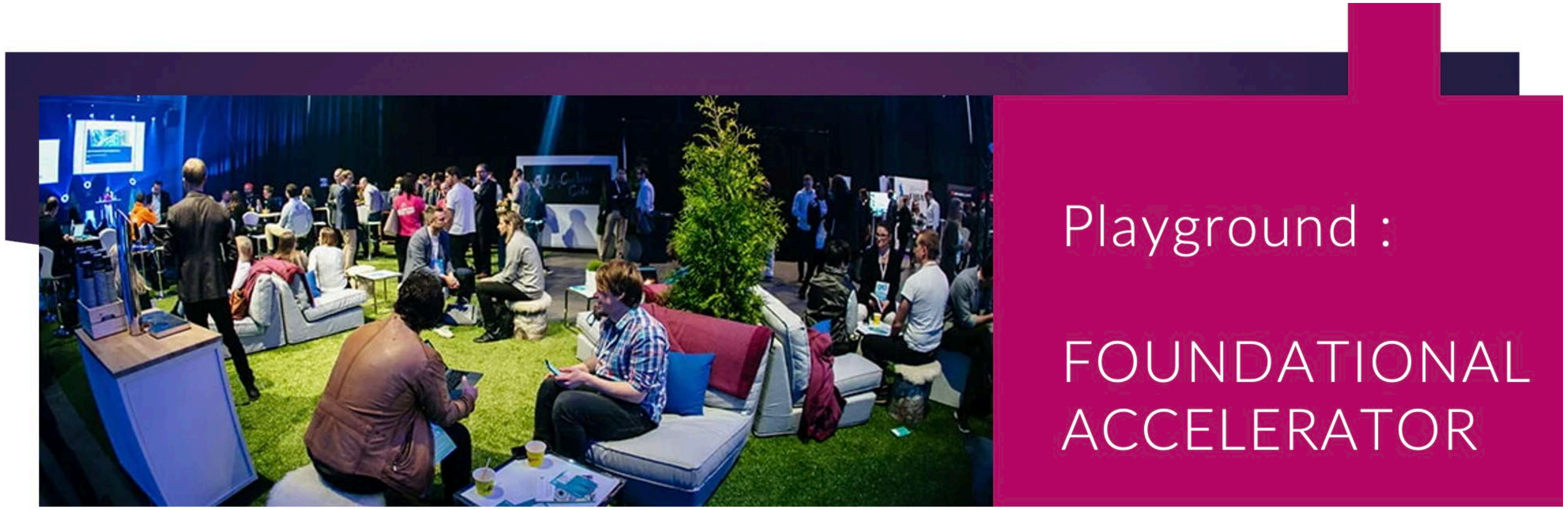
- GDPR : those who know it the best are GAFAM
- EU AI Act will be <u>key</u>

- Idea : AI & Law summer school (KU Leuven)

-> turn it into a broader event ?



Proposal 3 : add capital, foster startups



12 Corporate partners
(Engie, Suez, BNP,…)
12 Startups
(handpicked through application)
1 Demo Day
(with press, etc.)
150 K€ investment
(as convertible loan)

Best way to explore a new territory :

Define needs
Bring explorers
Make tools available
Provide funding

Proposal 3bis : storefront company

- Announcing Foundational.AI

Foundational AI

- as a service company around Foundation Models
- Extensions, fine-tuning, RLHF, …
- Teaming with large consulting companies
- Gathering expertise / talents / resources / visibility

(who wants to do a "split", JCVD-like ? ;-)

Wrap-up

- New paradigm
- Raises huge
  - Research questions
  - Challenges (legal, regulatory)
- Opens amazing opportunities
  - Academia / Regulatory
  - Startup(s)

+32 475 41 25 54
roald@roald.com
LinkedIn.com/in/roald

Foundational AI