“What AI Doesn’t Know”

(TACKLING THE BIG QUESTIONS)

Shane Ó Seasnáin Program Director AI

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Proposition #1: AI will kill us all

Joe Biden Grew More Worried About AI After Watching 'Mission: Impossible — Dead Reckoning,' Says White House Deputy

By Ethan Shanfeld

Google and Oxford scientists claim AI might destroy mankind in future

As per a recent research paper, advanced machines in the future will eventually become incentivised to break the rules their creators set and they would compete for limited resources or energy.

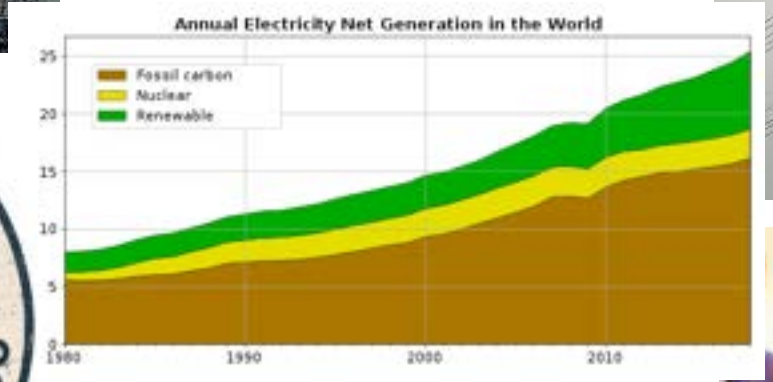
Rogue AI 'could kill everyone,' scientists warn as ChatGPT craze runs rampant

How Could A.I. Destroy Humanity?

Researchers and industry leaders have warned that A.I. could pose an existential risk to humanity. But they've been light on the details.

Five ways AI might destroy the world: 'Everyone on Earth could fall over dead in the same second'

Electricity will kill us all



Steel will kill us all

400 Cannon at Waterloo
3400 Artillery at Somme

Italian steel factory accused of thousands of deaths related to toxic emissions

The Truth about General-Purpose Technologies

People will use them to do harm and to do good

4 hours → 41 mins

Eliminate
360 to 1120 kt

In General, They Bring More Benefits

Proposition #2: AI will take our jobs

3 Perspectives

Wealth

Workers

Society & AI

Forget AI for a Moment: We need new ways of working

3 Perspectives

Wealth

Workers

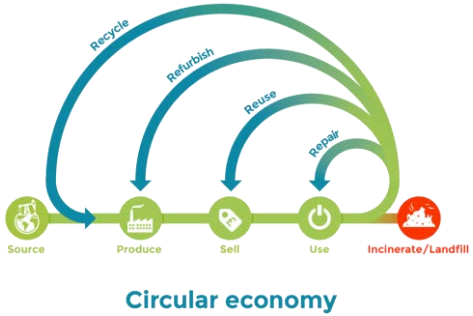
Society & AI

A better world with AI

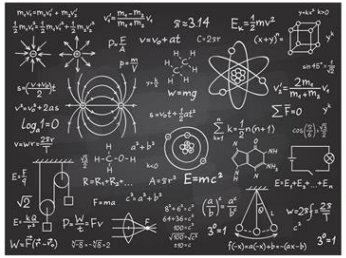
A New Opportunity to Re-frame Wealth & Work…

Economics
Healthcare
Energy
Raw Materials
Freedom

To make People Smarter
(Intelligence Augmentation)



6 Tribes of AI: Capturing Intelligence in a Computer

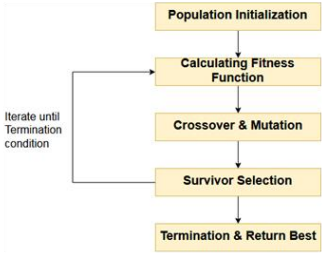


Logic

Symbolists

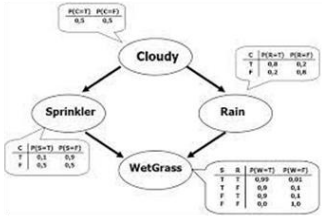
Patterns

Connectionists



Survival

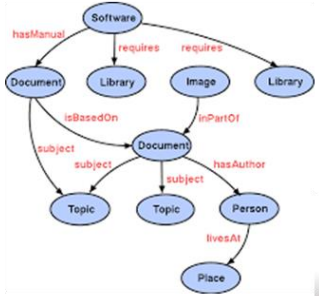
Evolutionists



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

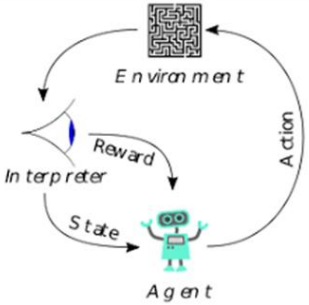
Probability

Bayesians



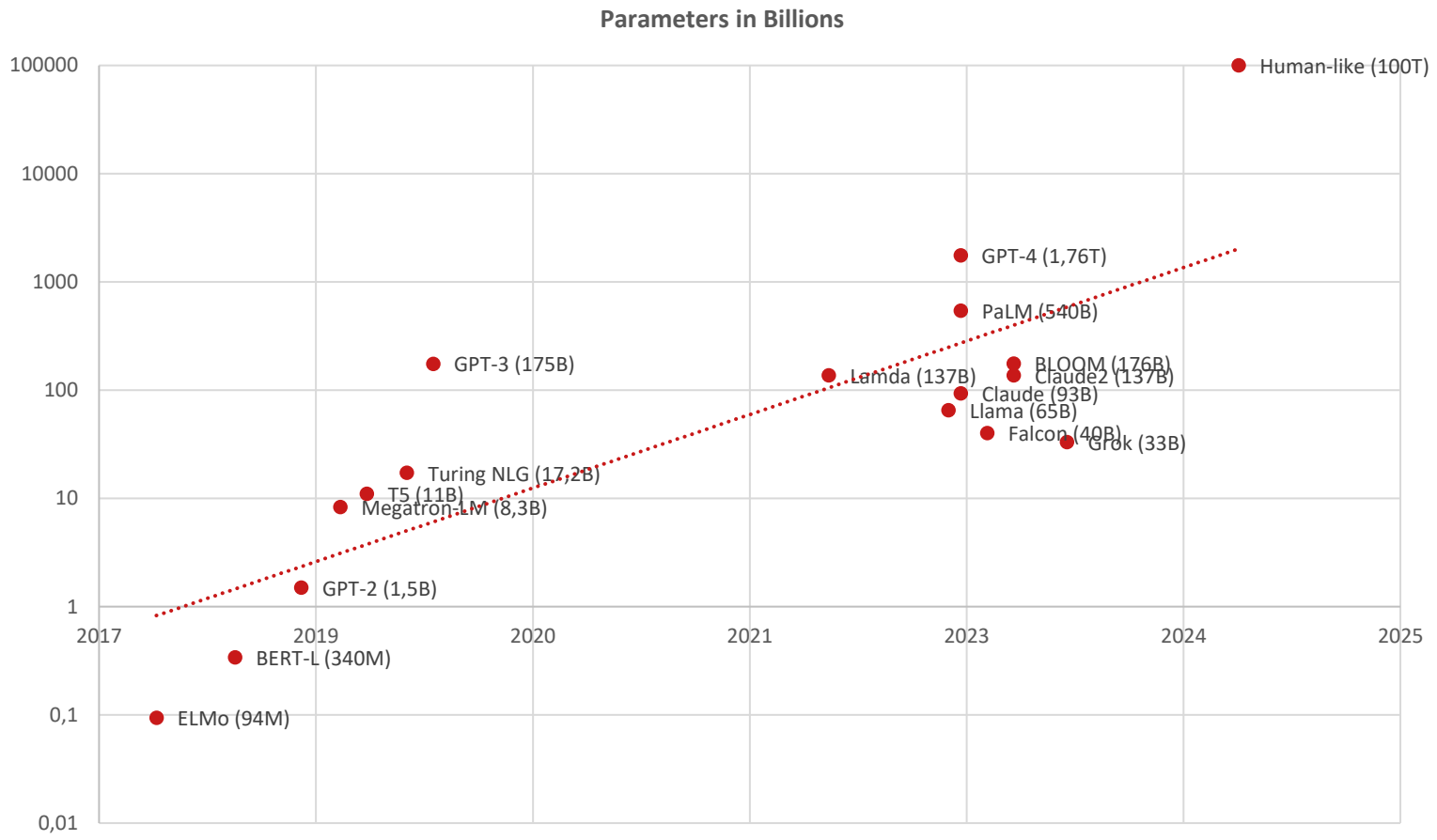
Similarity

Analogizers



Goals

Auto-didacts


Parameters in Billions
100000
10000
1000
100
10
1
0,1
0,01
2017
2019
2020
2021
2023
2024
2025
Human-like (100T)
GPT-4 (1,76T)
PaLM (540B)
GPT-3 (175B)
Lamda (137B)
BLOOM (176B)
Claude2 (137B)
Claude (93B)
Llama (65B)
Falcon (40B)
Grok (33B)
Turing NLG (17,2B)
T5 (11B)
Megatron-LM (8,3B)
GPT-2 (1,5B)
BERT-L (340M)
ELMo (94M)

Large Language Models

Web
Wiki
Corpus

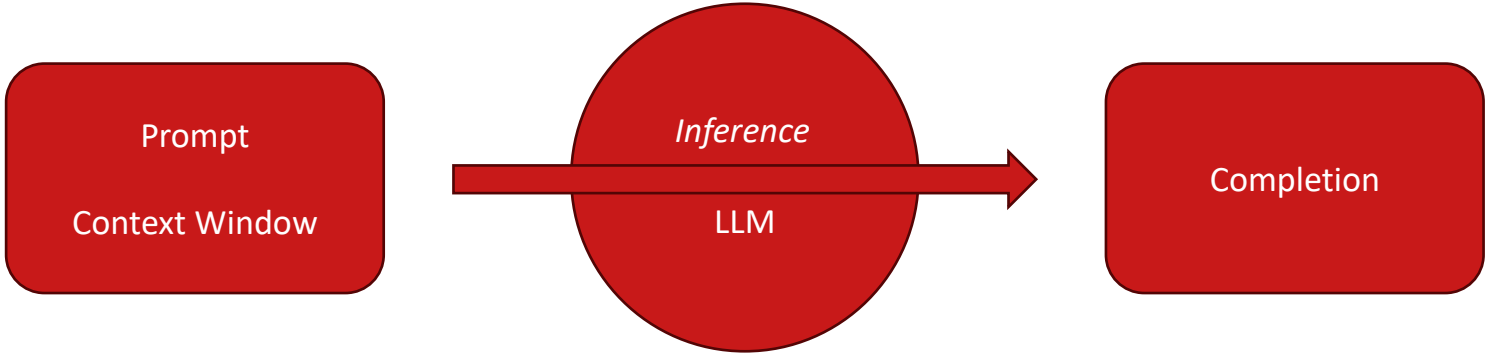
Training

LLM

Prompt Engineering

In-Context Learning:

- *Performed inside Context Window (limited size)*
- *Zero-shot inference = no examples*
- *One-shot inference = one example*
- *Few-shot inference = several examples (max 5-6)*



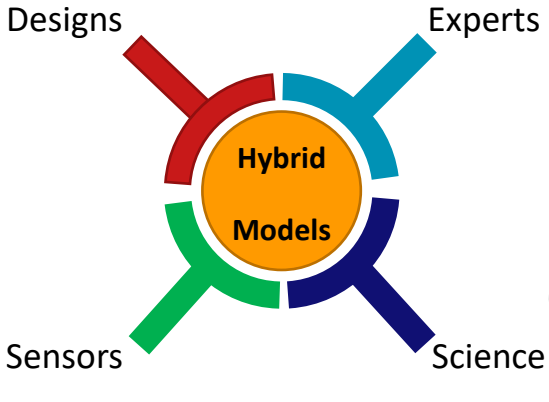
Instruction Fine-Tuning

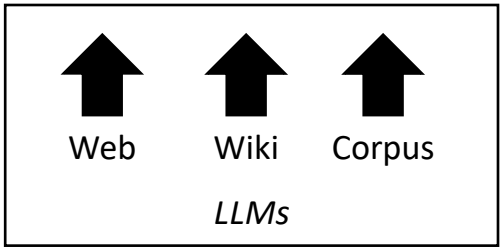
- Fine-Tuning on a Single Task can be done with 500 – 1k examples
- May however lead to **Catastrophic Forgetting** (forgetting how to do other tasks)
- Alternative: Parameter-Efficient Fine-Tuning (PEFT)

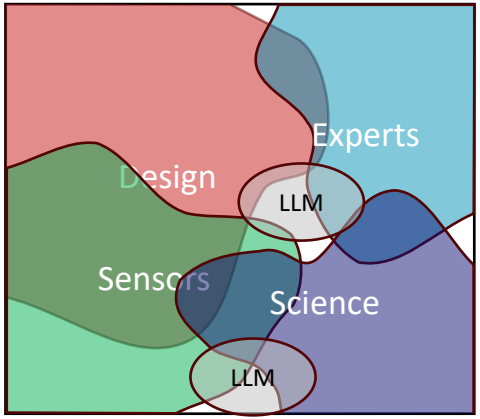
Problems with LLMs

Problems such as
- Out-of-Date Training Data
- Incorrect Responses (Maths)
- Hallucinations

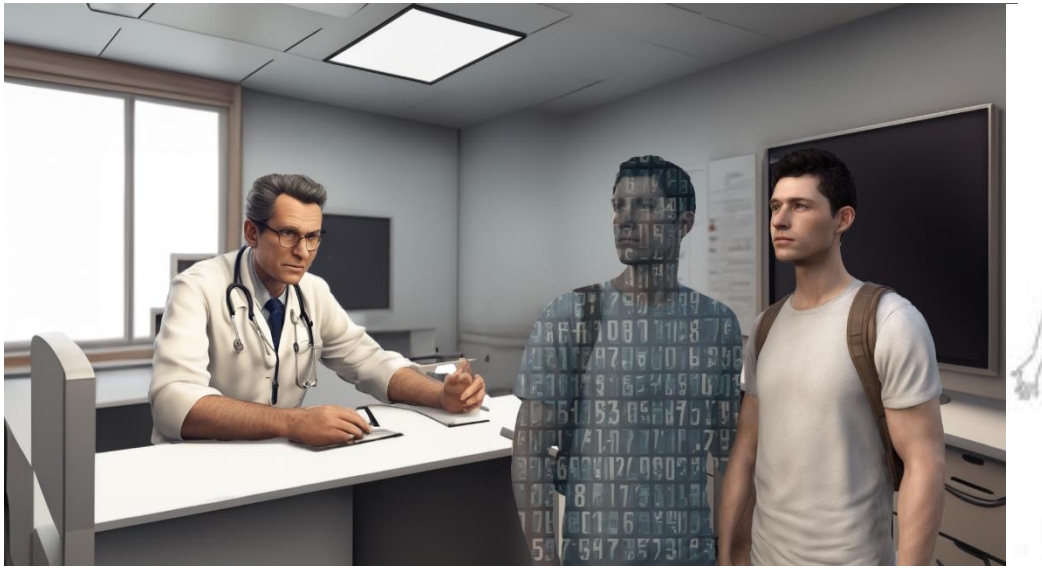
Problems in the Data

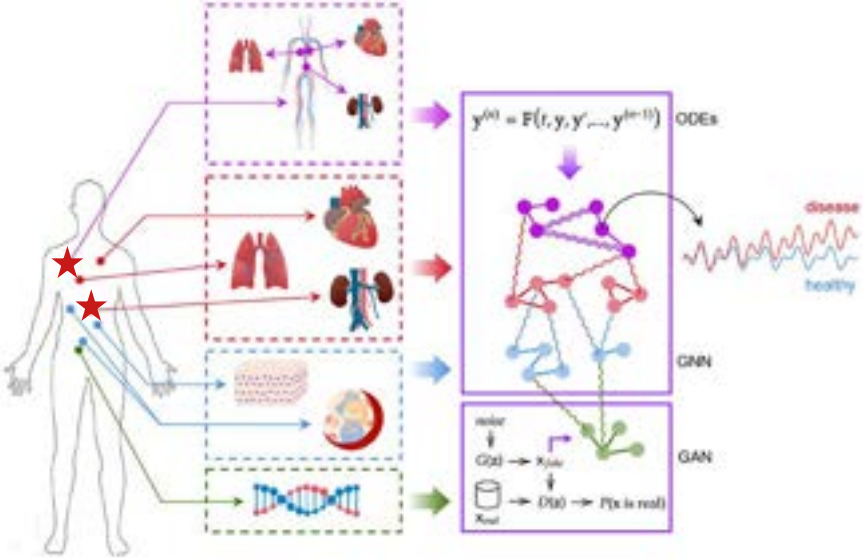






Augmenting Health Care





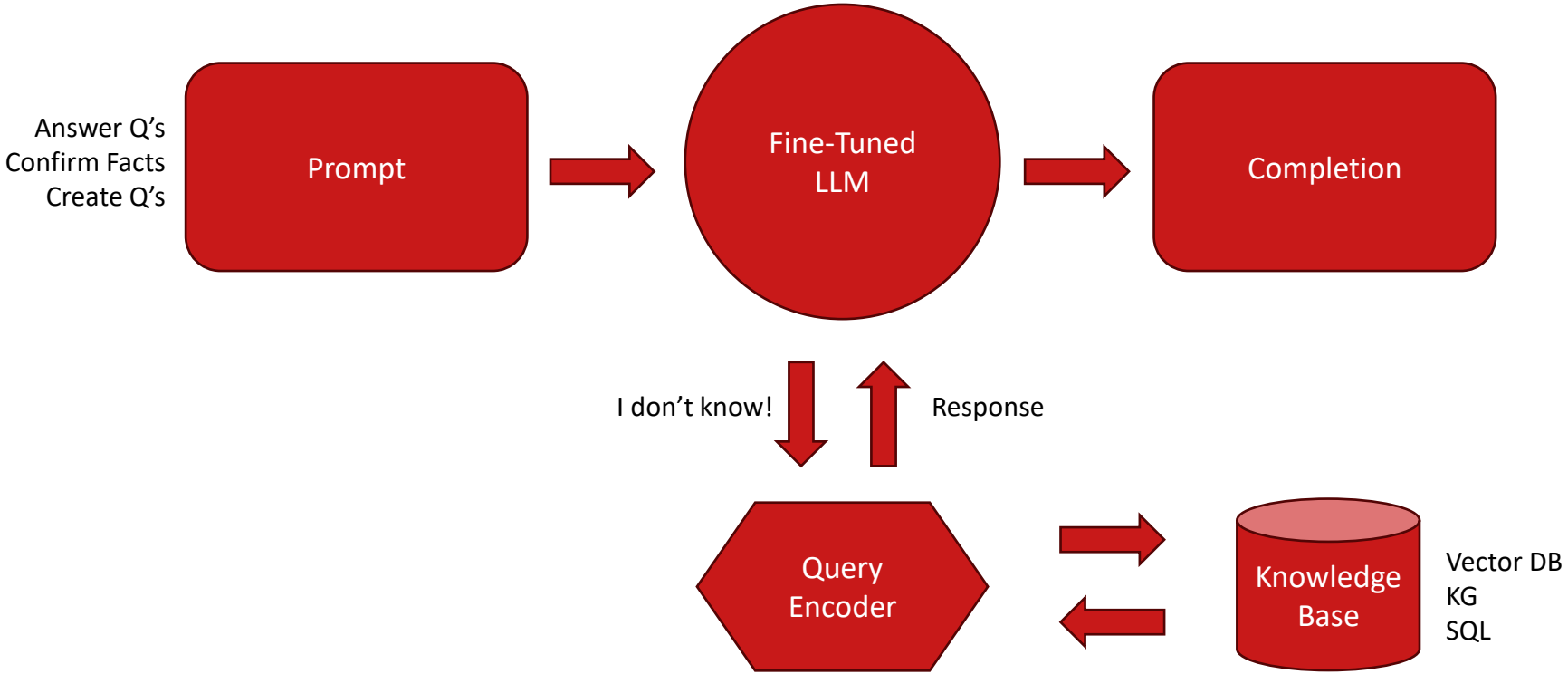
-5 Years

Now

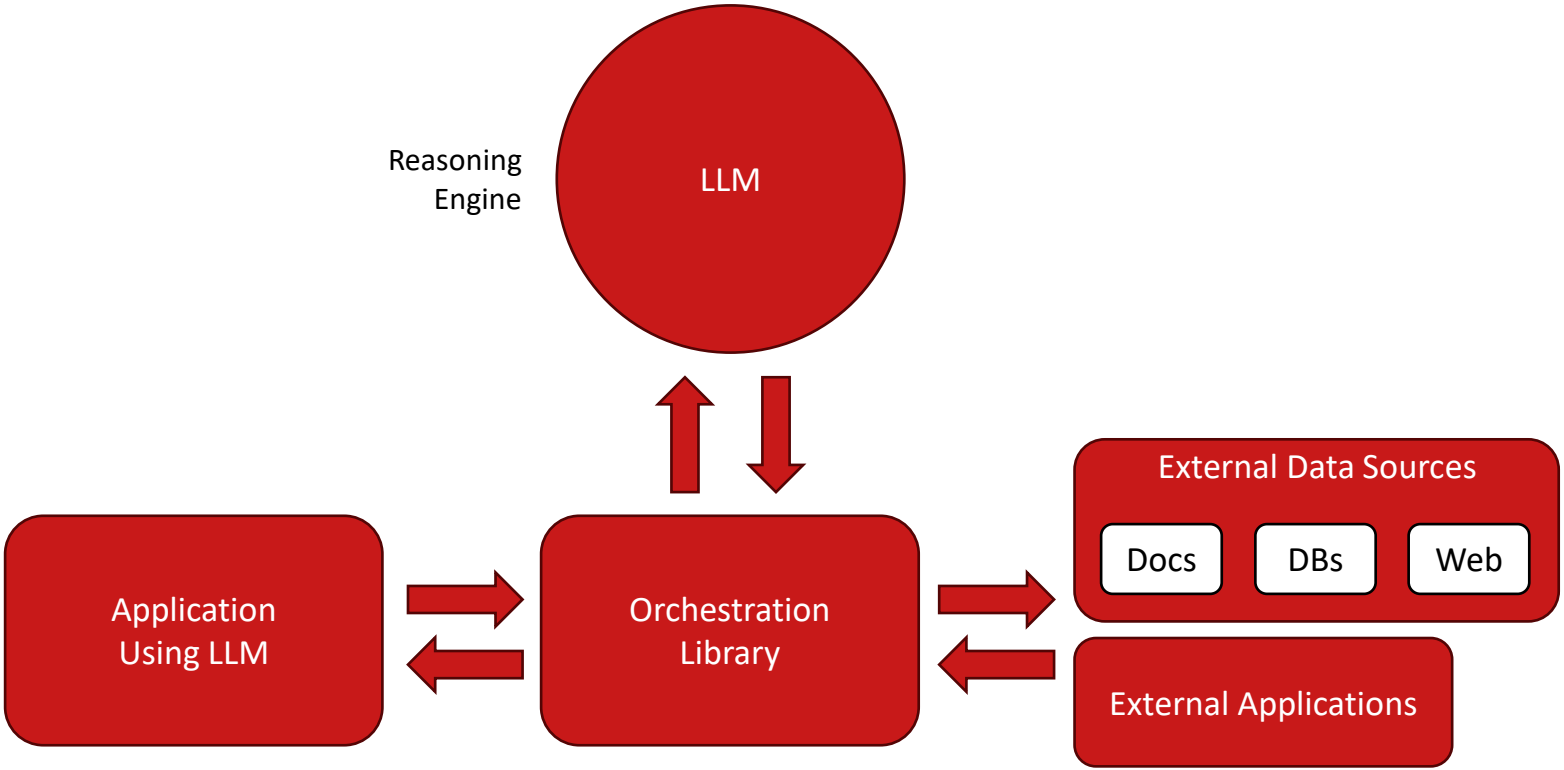
+5 Years

Augmenting Cities & Regions

Looking outside the LLM → Retrieval Augmentation Generation (RAG)



Looking outside the LLM → LLM Integration



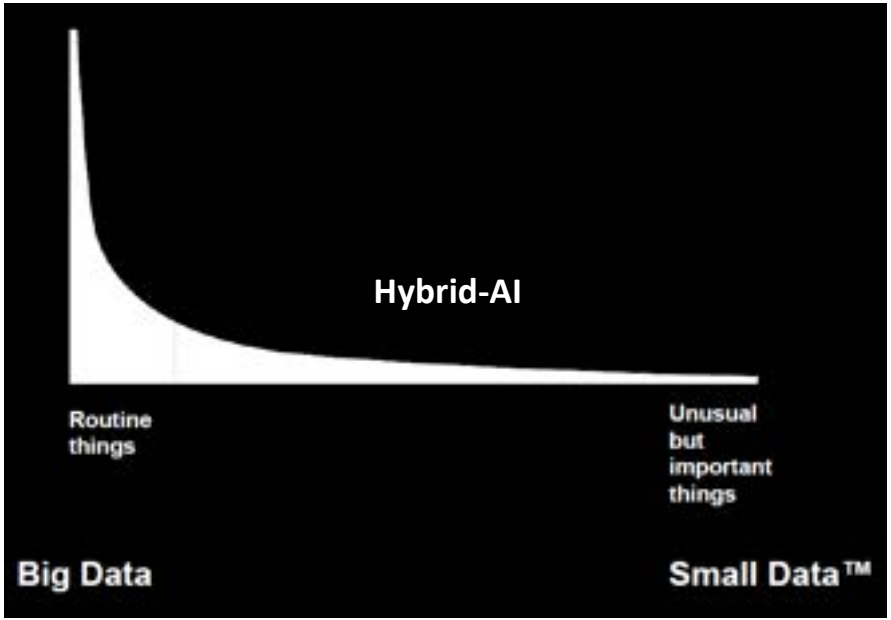
Augmenting Data

Make an appointment,
Translate text,
Image recognition

Personal advice,
Understand context,
Transparency

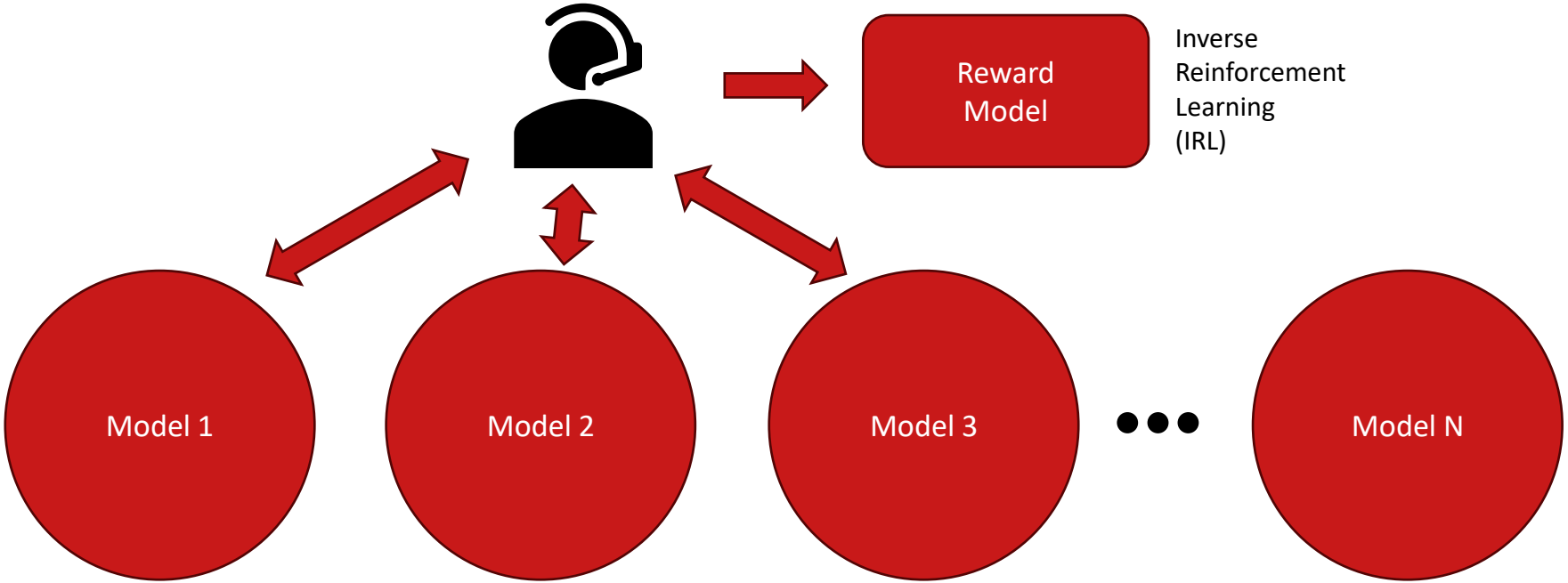
Deep-learning is strong

Deep-learning has challenges



Source: Gary Marcus @ GTC '21

Personal Goals & Alignment: MeGPT



Thank you!