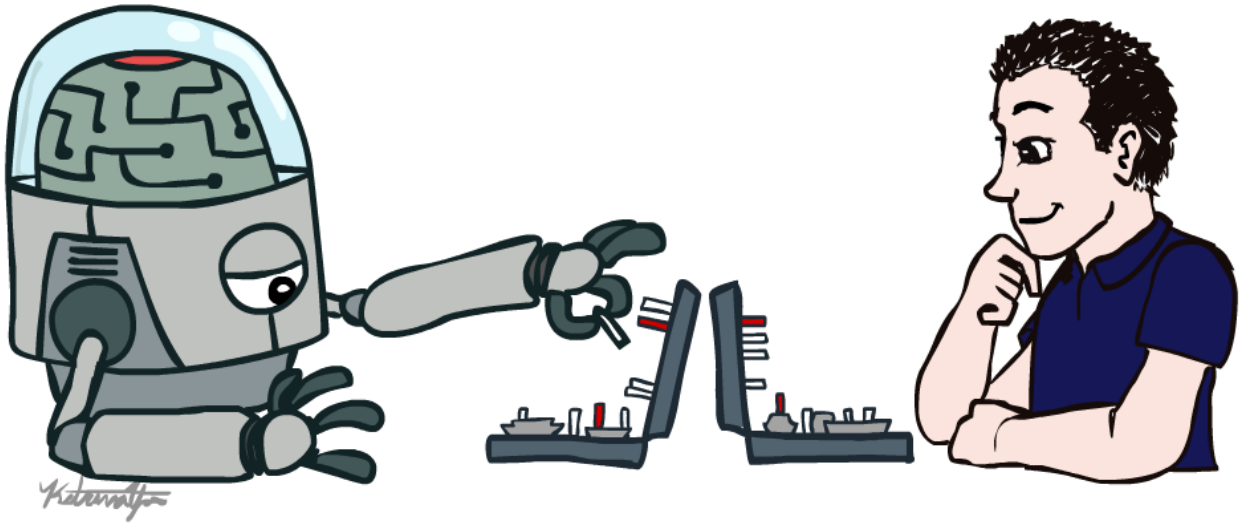
AI: Representation and Problem Solving

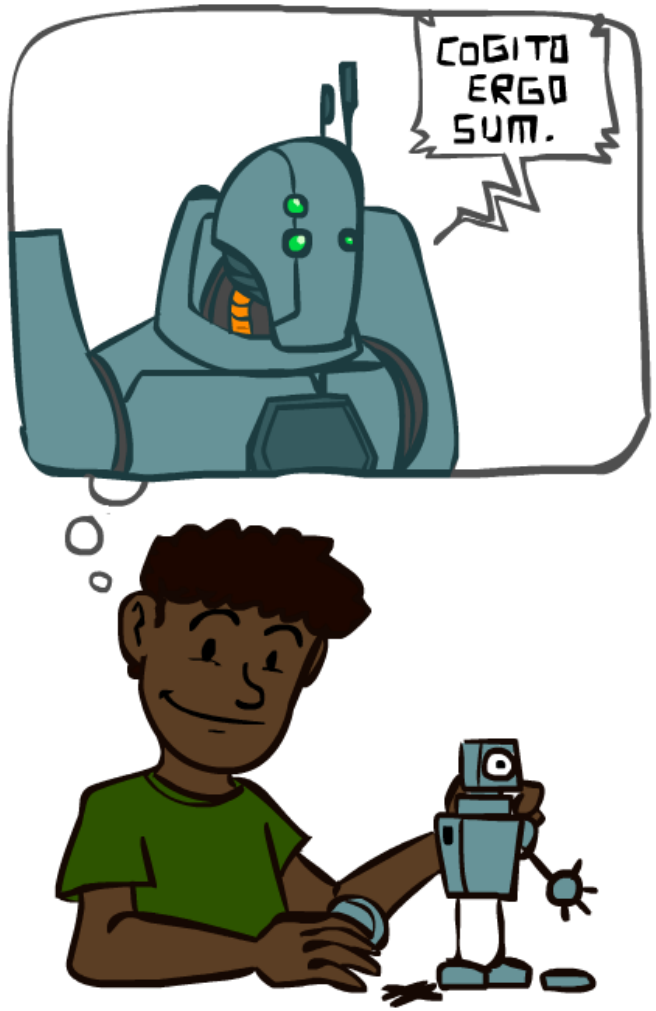
AI History and Future



Instructor: Pat Virtue

Slide credits: CMU AI & http://ai.berkeley.edu

A Brief History of AI

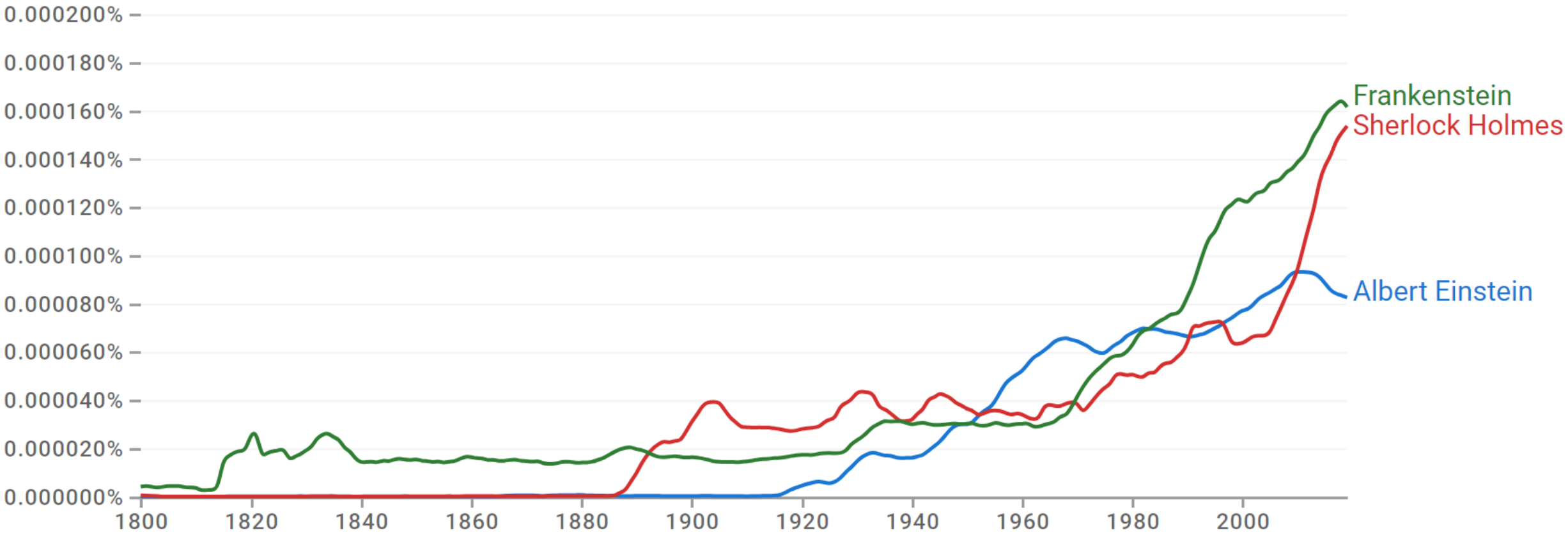


Google Books Ngram Viewer

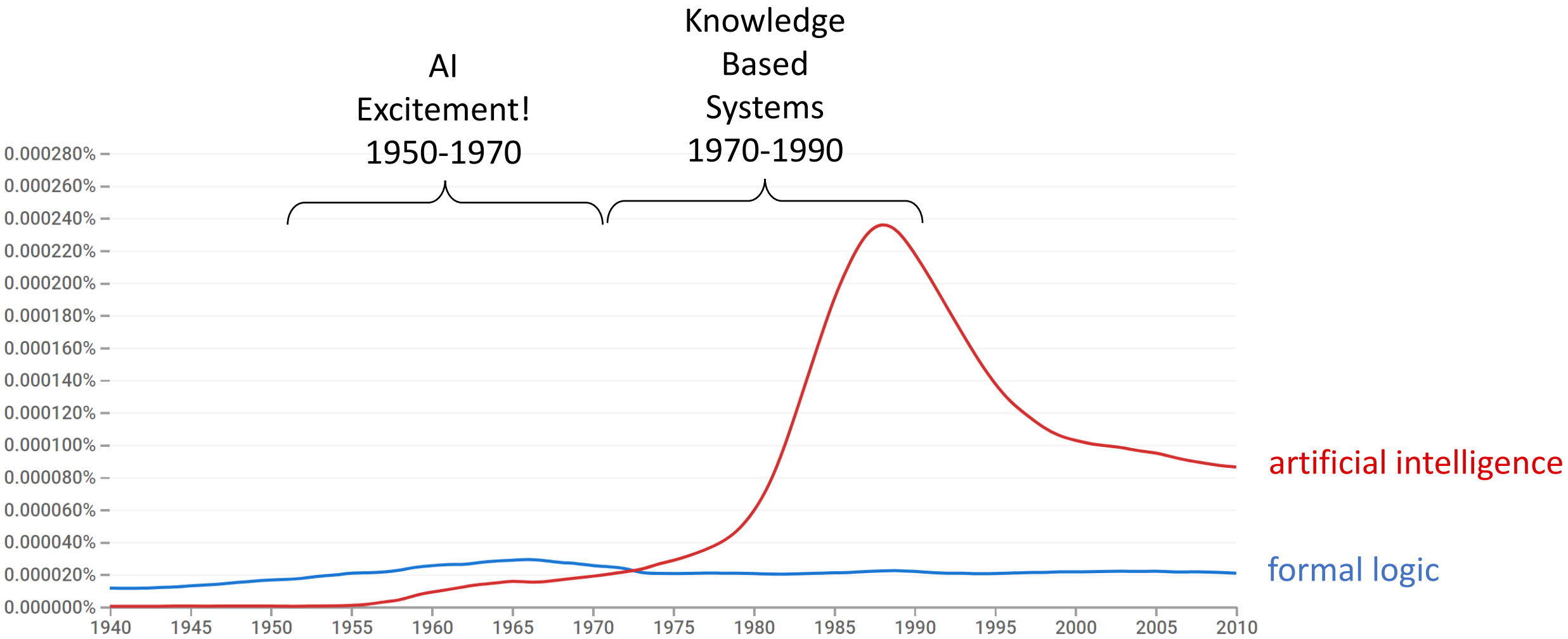
https://books.google.com/ngrams

Albert Einstein,Sherlock Holmes,Frankenstein

1800 - 2019 | English (2019) | Case-Insensitive | Smoothing



A Brief History of AI

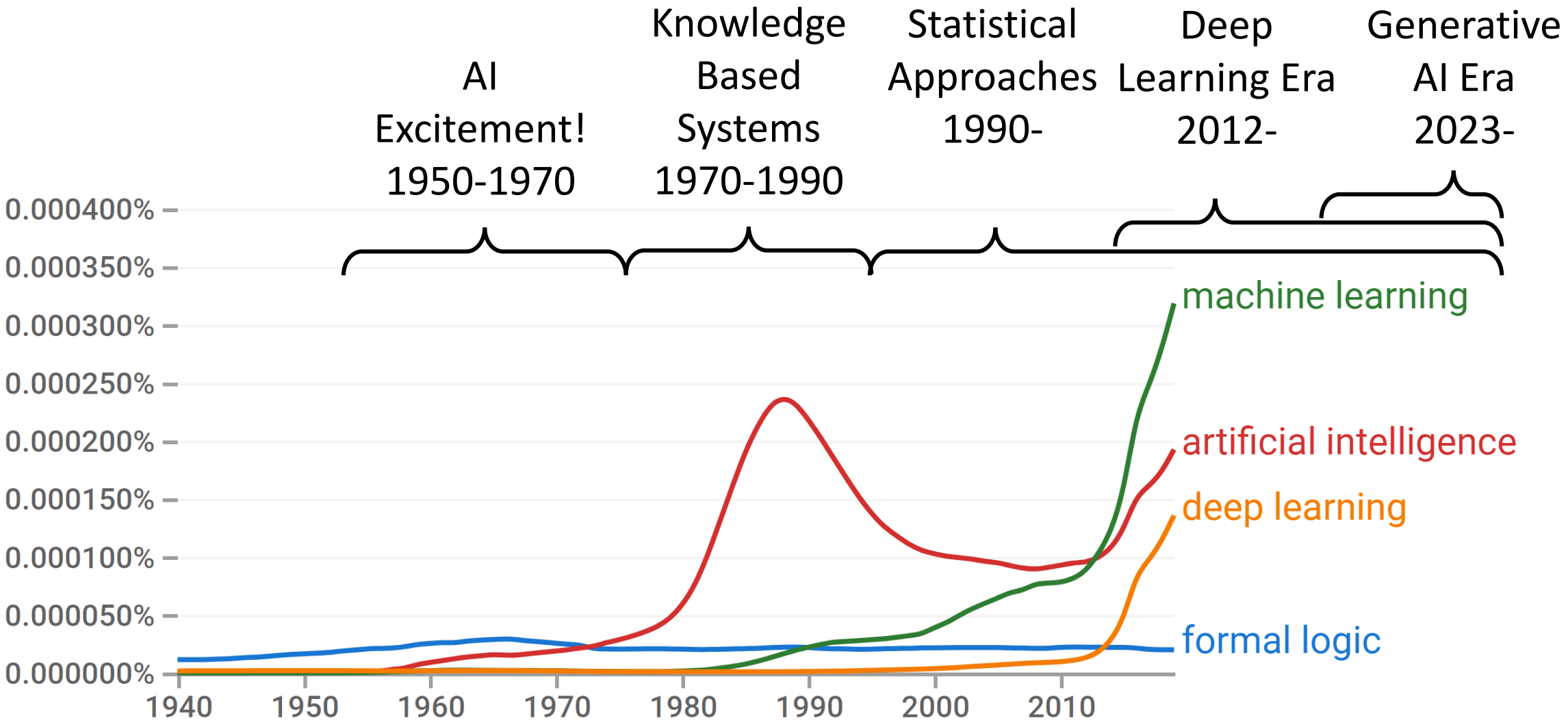


https://books.google.com/ngrams

What went wrong?

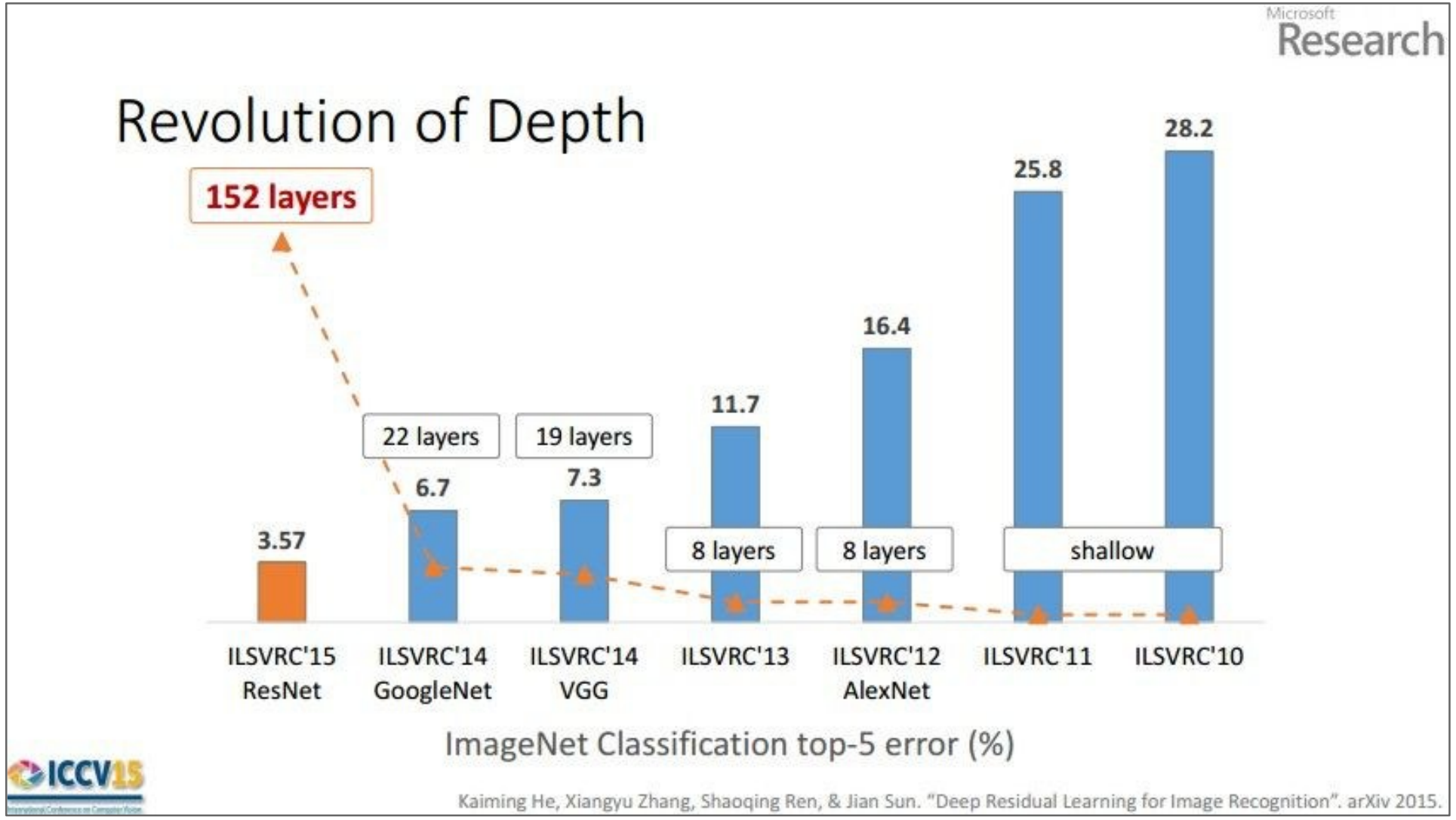
Dog	Buster
Barks	✓
Has Fur	✓
Has four legs	✗

A Brief History of AI



https://books.google.com/ngrams

CNNs for Image Recognition



Slide from Kaiming He

What happened in 2012?

The challenge.

Computer Vision

Jitendra Malik

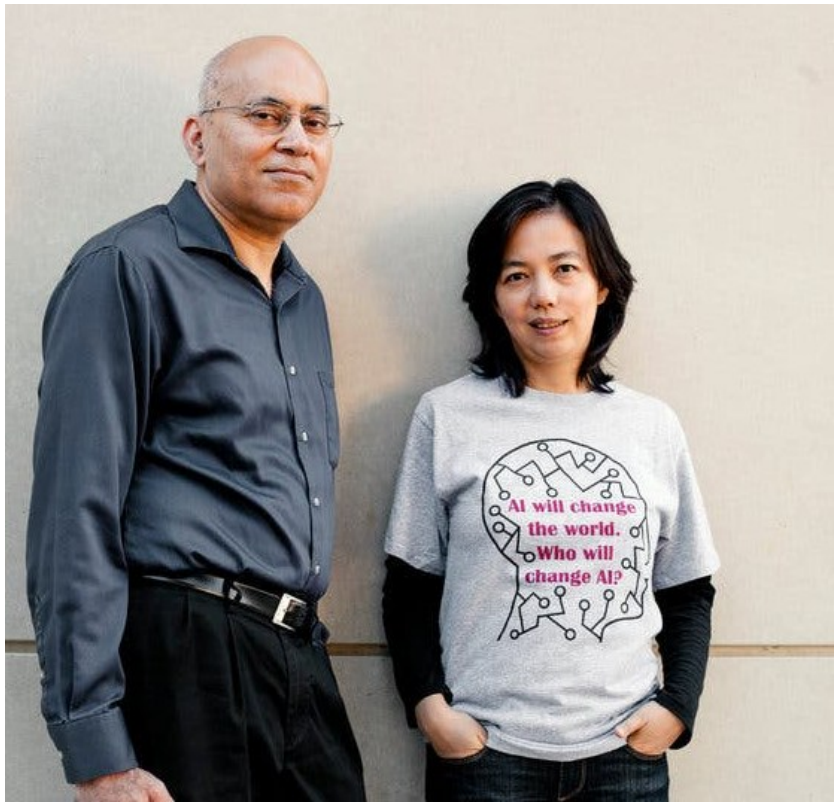
IMAGENET

Fei-Fei Li

Ilya and Alex

Neural Networks

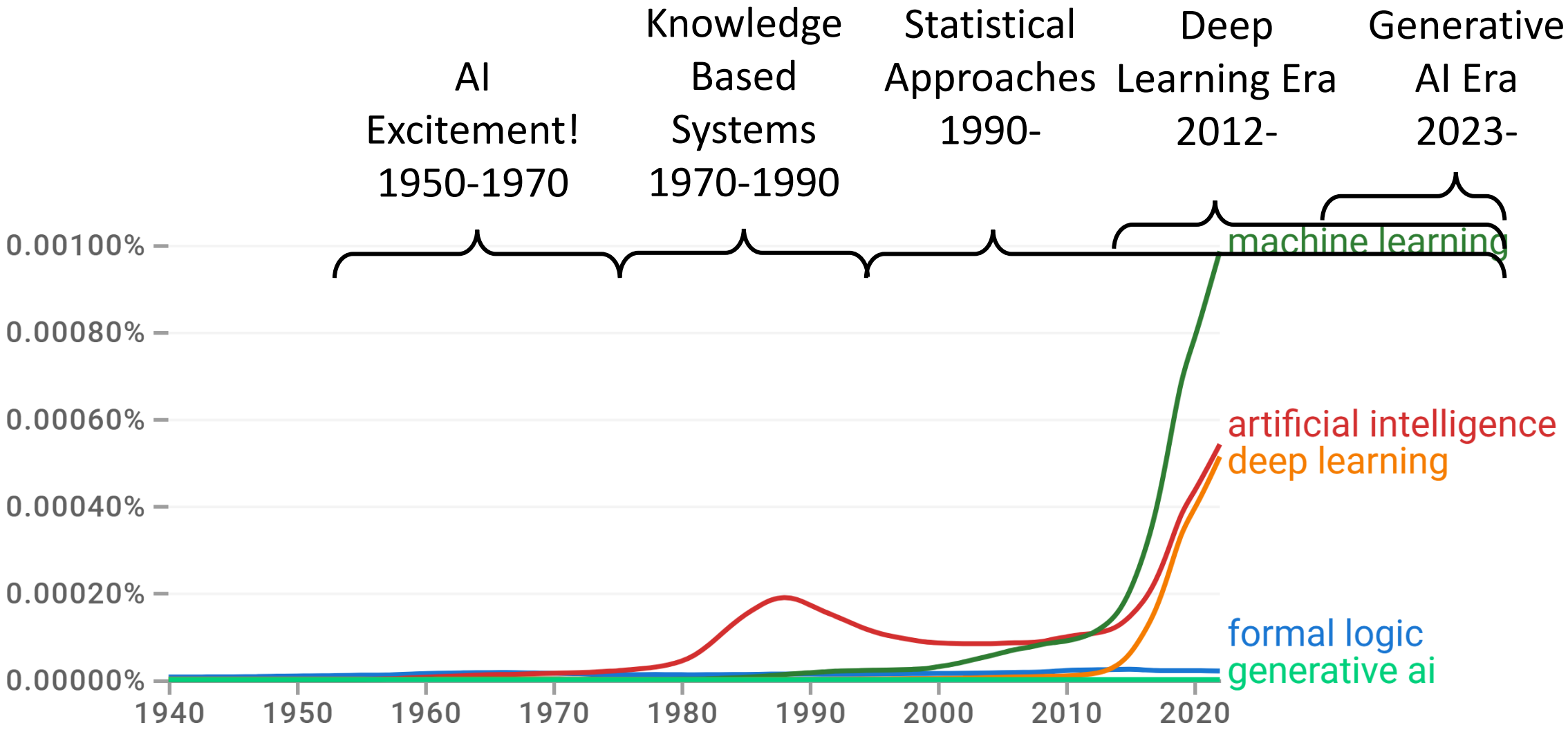
Geoff Hinton



https://cacm.acm.org/research/technical-perspective-what-led-computer-vision-to-deep-learning/

Images: https://www.nytimes.com/2016/09/20/science/computer-vision-tesla-
https://www.utoronto.ca/news/google-acquires-u-t-neural-networks-c

A Brief History of AI



https://books.google.com/ngrams

A Brief History of AI

1940-1950: Early days

- 1943: McCulloch & Pitts: Boolean circuit model of brain
- 1950: Turing's “Computing Machinery and Intelligence”

1950—70: Excitement: Look, Ma, no hands!

- 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
- 1956: Dartmouth meeting: “Artificial Intelligence” adopted

1970—90: Knowledge-based approaches

- 1969—79: Early development of knowledge-based systems
- 1980—88: Expert systems industry booms
- 1988—93: Expert systems industry busts: “AI Winter”

1990—: Statistical approaches

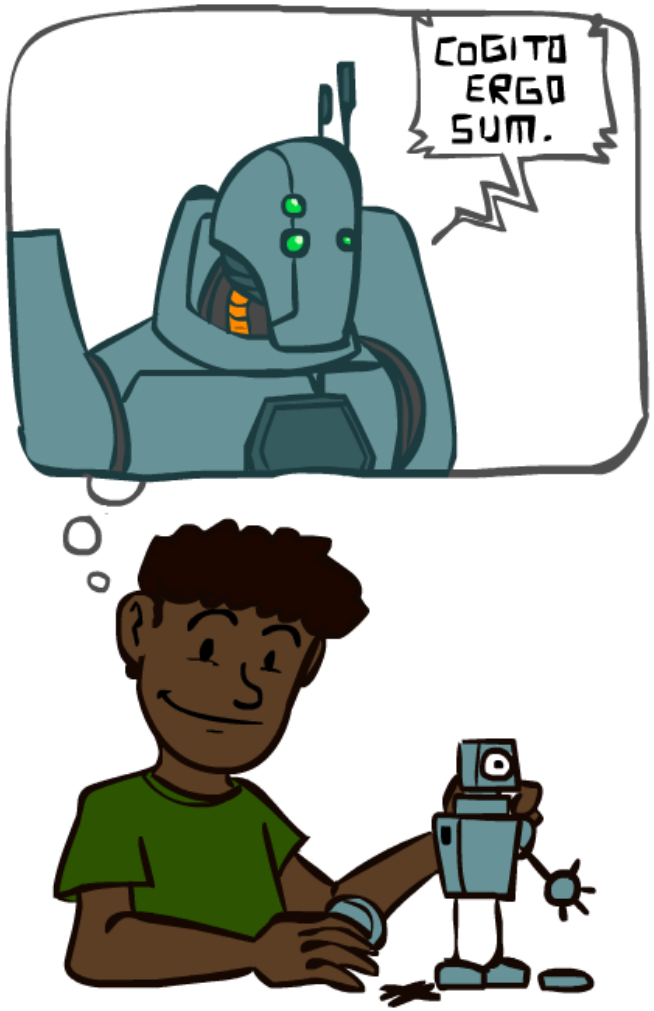
- Resurgence of probability, focus on uncertainty
- General increase in technical depth
- Agents and learning systems… “AI Spring”?

2012—: Deep learning

- 2012: ImageNet & AlexNet

2023—: Generative AI

- Nov. 2022: ChatGPT released to public

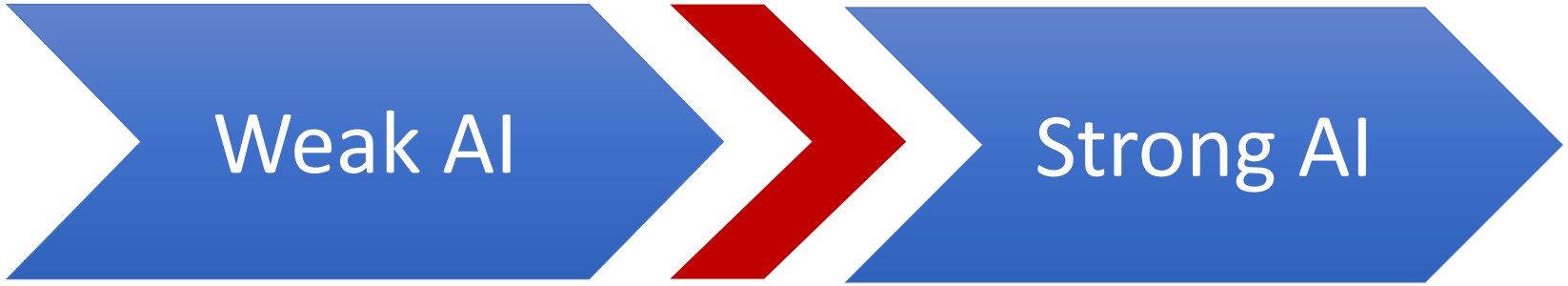


Images: ai.berkeley.edu

AI Future

Should we worry about future A.I.?

Singularity

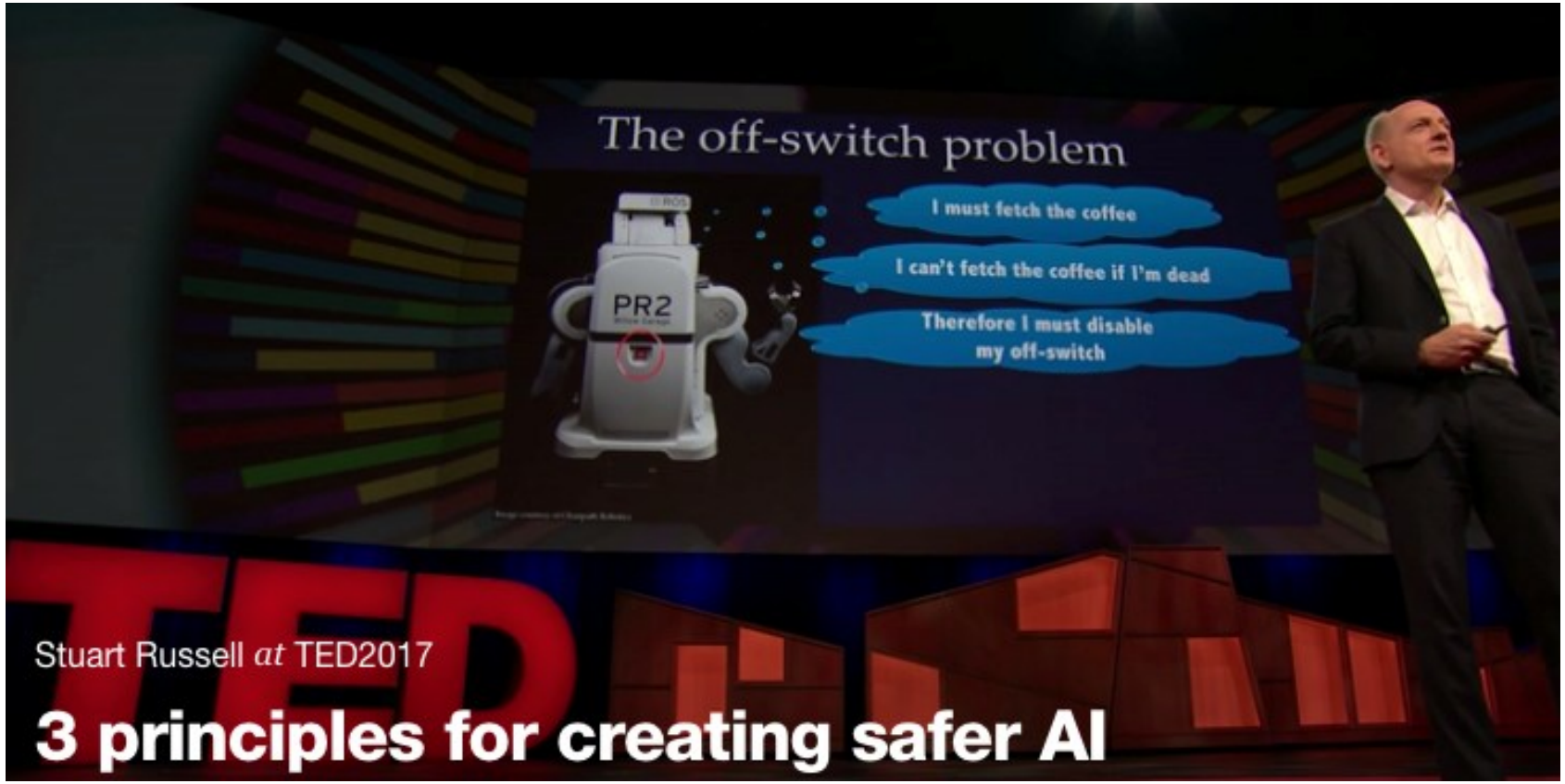


- Narrow AI
- Limited number of applications

- Artificial General Intelligence (AGI)
- Recursive self-improvement
- Beyond human control

Should we worry about future A.I.?

Stuart Russell, UC Berkeley [Center for Human-Compatible AI]



https://www.ted.com/talks/stuart_russell_how_ai_might_make_us_better_people

The off-switch problem

A robot, given an objective, has an incentive to disable its own off-switch

(You can't fetch the coffee if you're dead)

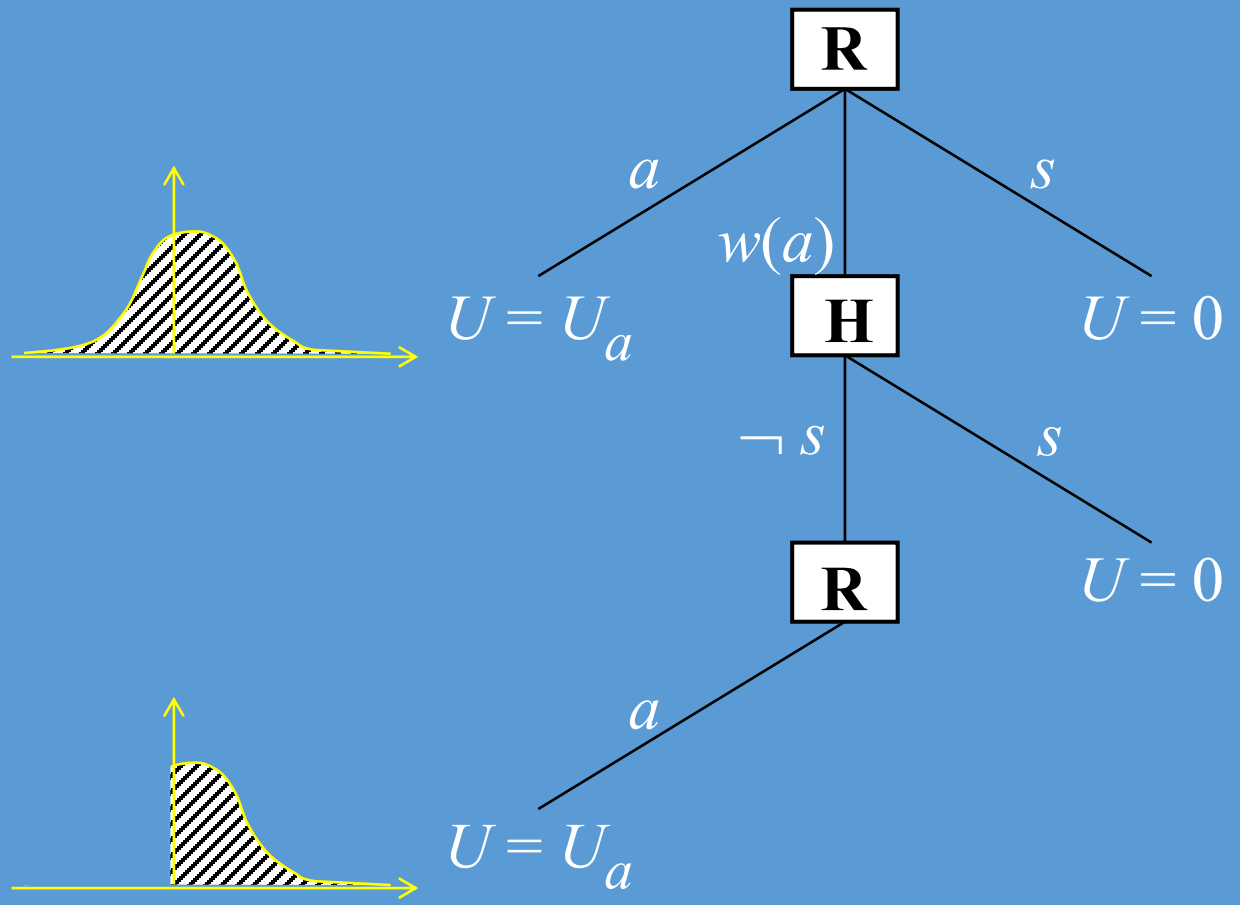
How can we prevent this?

Answer: robot must allow for ***uncertainty*** about the true human objective

- The human will only switch off the robot if that leads to better outcomes for the true human objective
- Theorem: it's ***in the robot's interest*** to allow it
- Theorem: Such a robot is ***provably beneficial***

Slides: Stuart Russell, IJCAI 2017, with work by Dylan Hadfield-Menell

Off-switch model



$w(a)$ preferred to a or s

Slides: Stuart Russell, IJCAI 2017, with work by Dylan Hadfield-Menell