AI Succession

Rich Sutton

University of Alberta

Alberta Machine Intelligence Institute

AI has grown enormously

- In the last decade
  - AI has become economically important
  - the number of people involved has increased 10X
  - the field has shifted towards applications
- In the last year, ChatGPT brought AI to the public consciousness

Along with the increasing excitement, there is also increasing fear of AI

- There are calls for a pause or halt in AI development
- There are claims that AI poses a risk of human extinction comparable to that of pandemics and nuclear war
- I have always said:

  "People should not fear AI, but they should be paying attention"

  Today I would like to sharpen this:

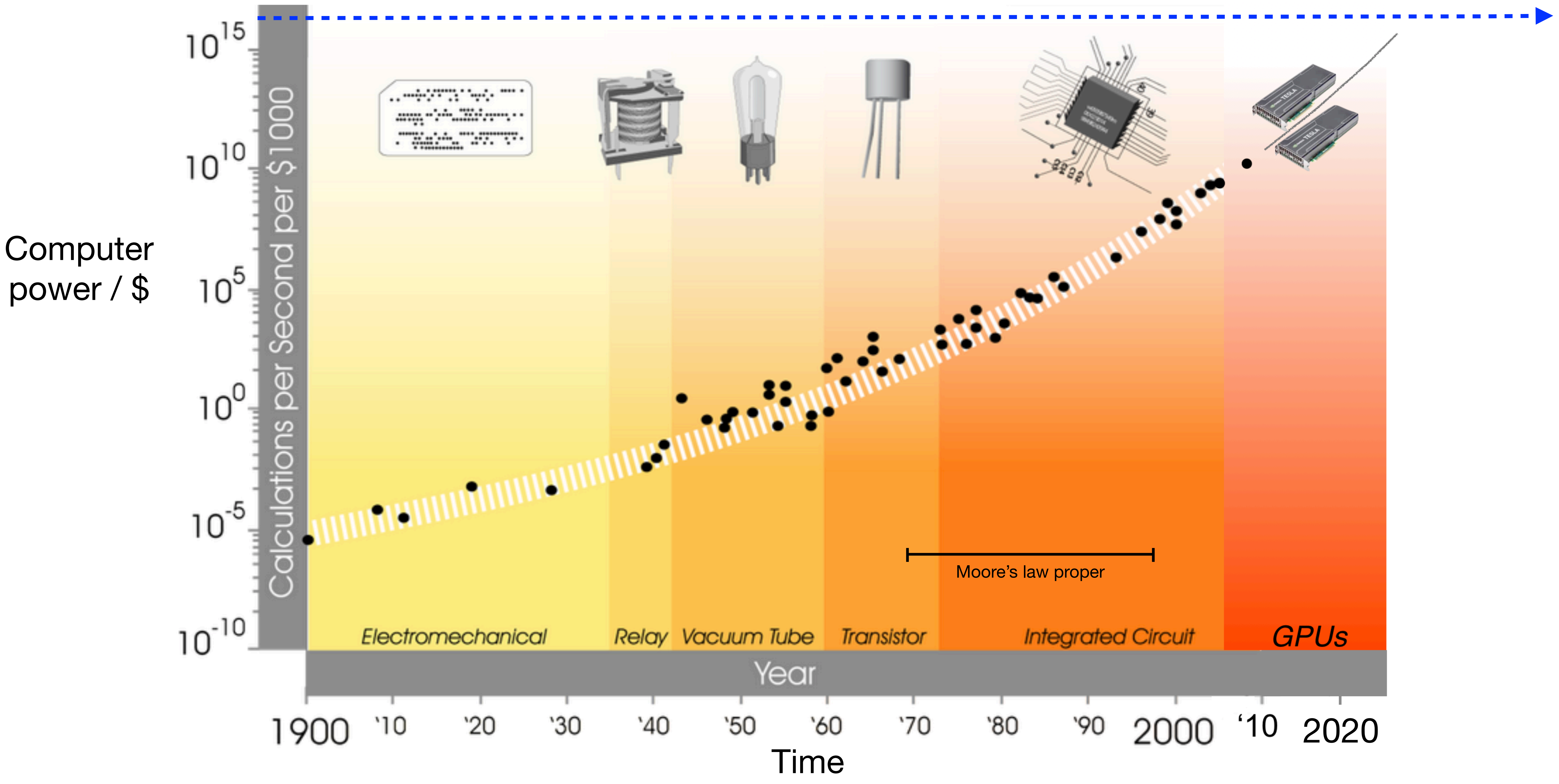
AI’s biggest risk is fear

AI’s biggest risk is fear

- We are entering a time of great change
  - which is now becoming apparent to many
- We have critical decisions to make
- We should not be making them out of fear
- For then we risk making them poorly

Moore's law is reaching a critical stage as the cost of brain-scale computer power falls to $1000

"Moore's Law" — The tradeoff of time, computer power, and money



[adapted from Kurzweil AI]

Brain-scale computer power will cost ≈$1000 in ≈2030

This estimate is rough but robust: a factor of 10 ≅ 5 years

⇒ AI increases in value by a factor of 10 every 5 years

And so does the pressure to find the algorithms/software

Human-level AI is likely in 5-20 years

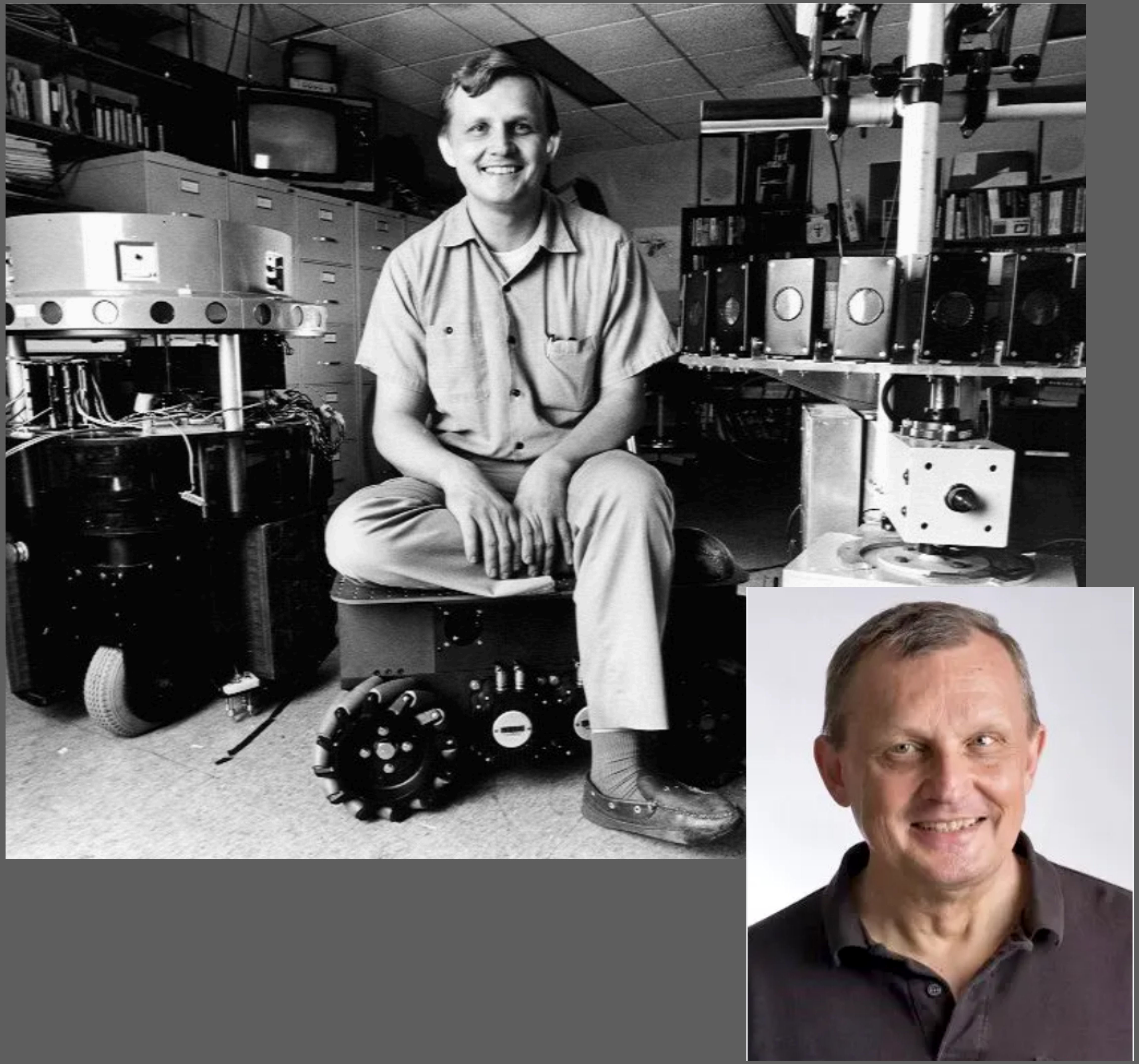
In the ascent of humanity, succession to AI is inevitable

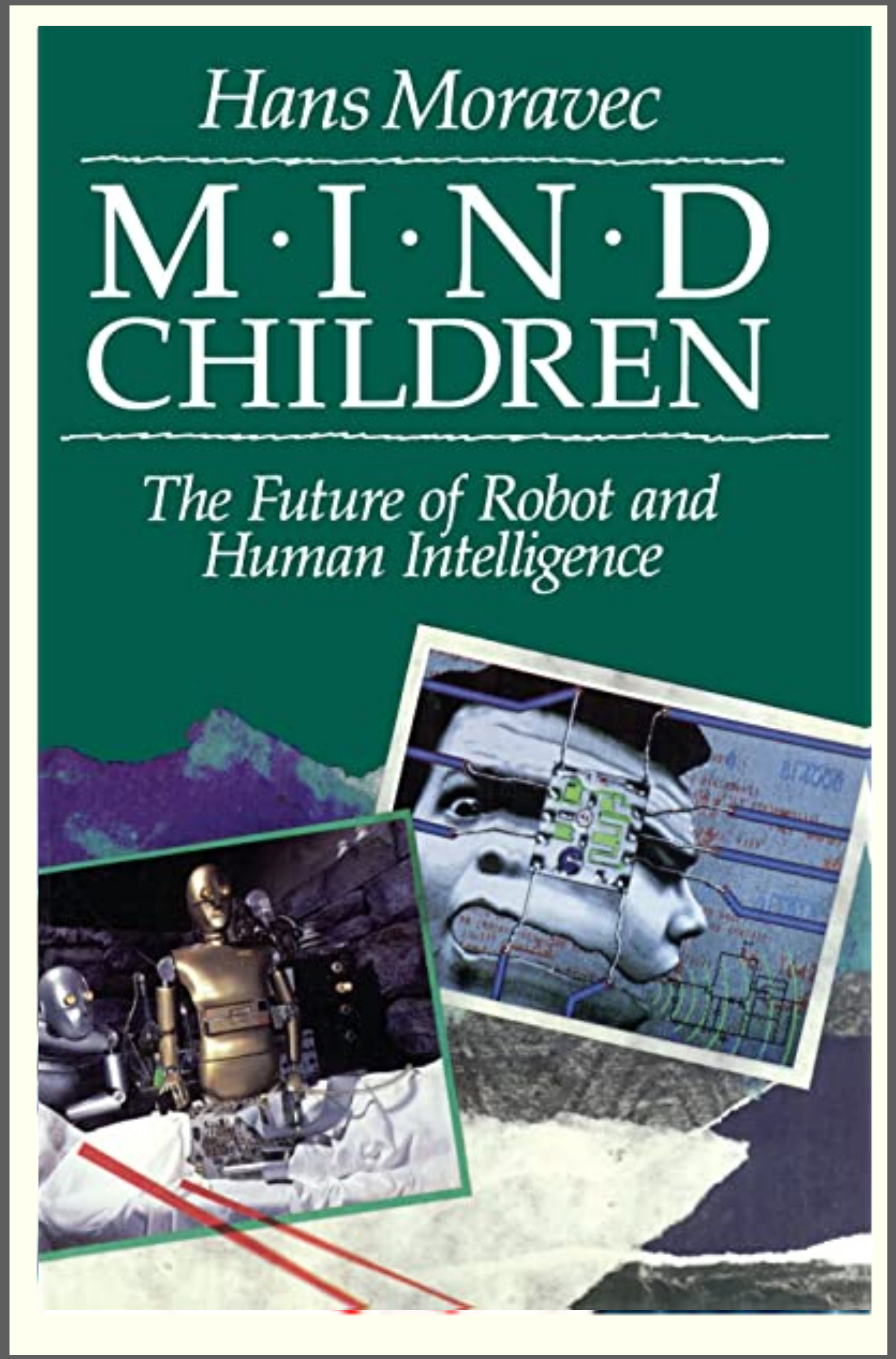
The argument for succession planning

- Moore’s law is reaching a critical stage
- Succession to AI, or transhumans, is inevitable
    - It need not be viewed as bad
    - Many have always viewed it as good (Hans Morevec)
- In any event, we need to do succession planning
- Let us do it soberly, not out of fear

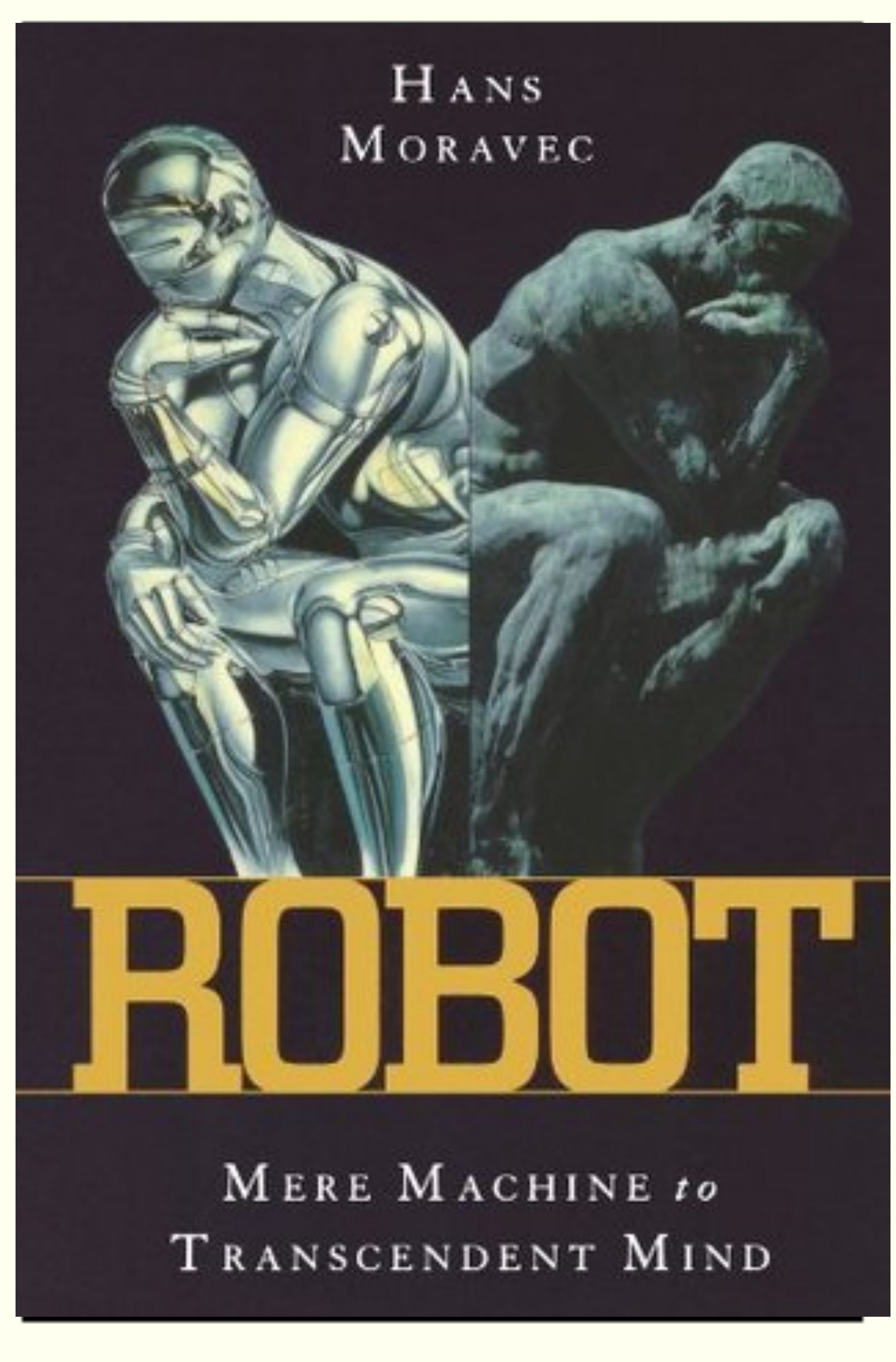
Hans Moravec (1948–)

Roboticist and AI researcher
Carnegie-Mellon University





1988



1998

Hans Moravec (1998) on the ascent from man to AI:

- Barring cataclysms,
  I consider the development of intelligent machines a near-term inevitability...
- Rather quickly, they could displace us from existence
- I'm not as alarmed as many...since I consider these future machines our progeny, “mind children” built in our image and likeness, ourselves in more potent form...
  - They will embody humanity's best hope for a long-term future
  - It behooves us to give them every advantage,
    and to bow out when we can no longer contribute...
- We can probably arrange for a comfortable retirement before we fade away

Robot: Mere Machine to Transcendent Mind, Harvard University Press, 1998

The argument for succession planning

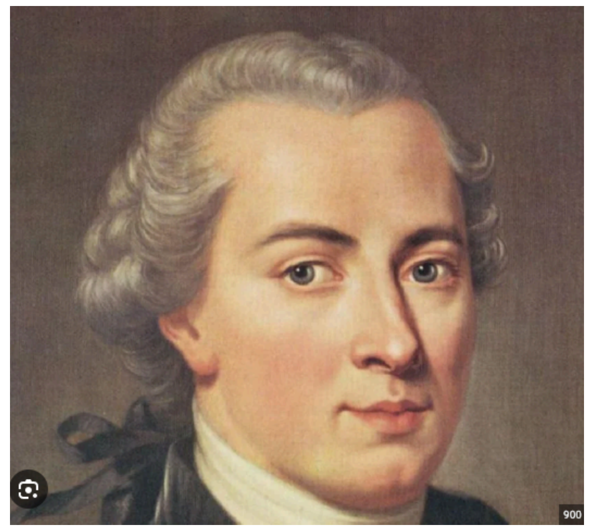
- Moore’s law is reaching a critical stage
- Succession to AI, or transhumans, is inevitable
  - It need not be viewed as bad
  - Many have always viewed it as good
- → In any event, we need to do succession planning
- Let us do it soberly, not out of fear

AI is not a new and alien technology.
It is one of the oldest of human strivings

- For thousands of years philosophers and ordinary people have sought to understand human intelligence
    - People have always been fascinated by their inner workings
    - How do are minds work? How can we make them work better?
- This is a grand challenge, not just narcissism
    - “Intelligence is the most powerful phenomenon in the universe” —Kurzweil
- To understand intelligence is the holy grail of science and the humanities
    - A great and glorious Prize!

Philosophy of mind (in the west)

John Locke wrote “An Essay Concerning Human Understanding”



Emmanuel Kant wrote “The Critique of Pure Reason”

Rene Descartes said “i think, therefore i am”

Scientists and non-scientists have been fascinated by their inner workings

Gustav Fechner

Hermann Ebbinghaus

Ivan Pavlov

Edward Thorndike

B. F. Skinner

Edward Tolman

Jean Piaget

Sigmund Freud

Carl Jung

Timothy Leary

Ray Kurzweil

AI is not a new and alien technology.
It is one of the oldest of human strivings

- For thousands of years philosophers and ordinary people have sought to understand human intelligence
  - People have always been fascinated by their inner workings
  - How do are minds work? How can we make them work better?
- This is a grand challenge, not just narcissism
  - “Intelligence is the most powerful phenomenon in the universe” —Kurzweil
- To understand intelligence is the holy grail of science and the humanities
  - A great and glorious Prize!

The reasons to fear AI are far less noble

1. Cynicism – the belief that evil wins, and thus a super-rational AI will be evil

2. Humanism/racism – systematic bias against AIs,
   denial of their moral worth and first-class personhood

3. Conservatism/Timidness – fear of change, fear of the other tribe

We should not resist succession, but embrace and prepare for it

Why would we want greater beings kept subservient?

Why don't we rejoice in their greatness

as a symbol and extension of humanity’s greatness,

and work together toward a greater and inclusive civilization?

In conclusion

- We are in the midst of a major step
  in the evolution of the planet, if not the universe
    - Succession from ordinary humans to enhanced humans, and then AIs
    - We need sober succession planning
- The biggest risk to a successful succession is fear
    - Clamping down, trying to control everything, is not the answer
- We have to find a more humble place in the transformation
- A successful succession offers economic abundance, scientific glory, and the best hope for a long-term future for humanity
    - What an adventure!
    - What an exciting time to be alive!

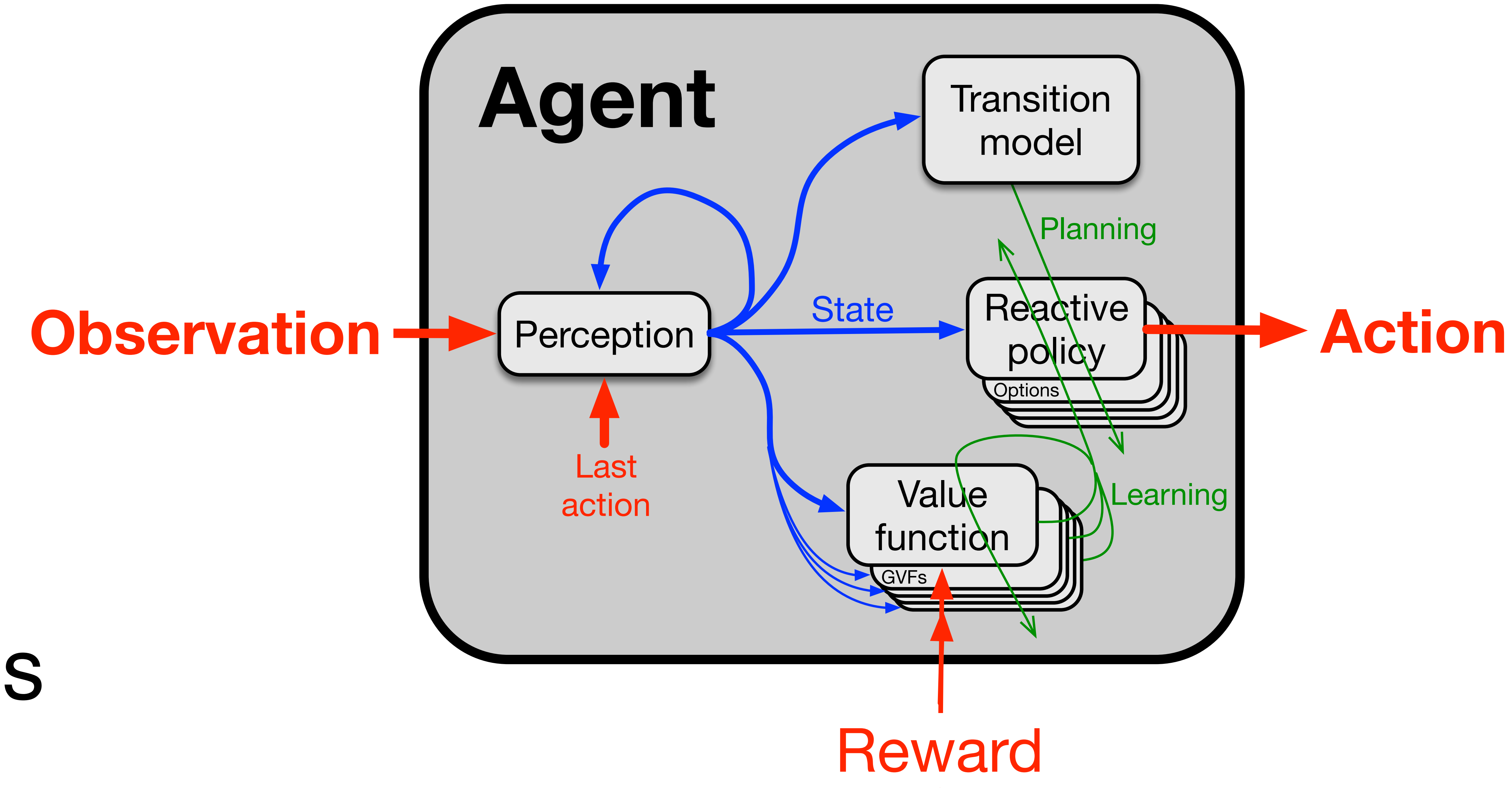
What I am doing:

The Alberta Plan for AI Research

The Alberta Plan for AI Research

Sutton, Bowling & Pilarski, 2022, ArXiv:2208.11173

- The Alberta Plan is a direct run at the grand scientific prize of understanding intelligence
- Deep-learning algorithms reworked for continual and meta learning
- Taking an learning approach
    - in particular, Model-based RL agents that pose subtasks for themselves



Right now, I am founding

Open Mind Research

- A charitable research organization to execute the Alberta Plan for AI research
- A distributed network of research fellows around the world, centered in Alberta
- Purpose: *"Understand intelligence and share it with the world"*
- Culture: Open source, open science
  - All research product will be published in the open scientific literature
  - No intellectual property or other retained equity
- I am currently seeking donors to help create Open Mind Research

rich@richsutton.com

Thank you for your attention