

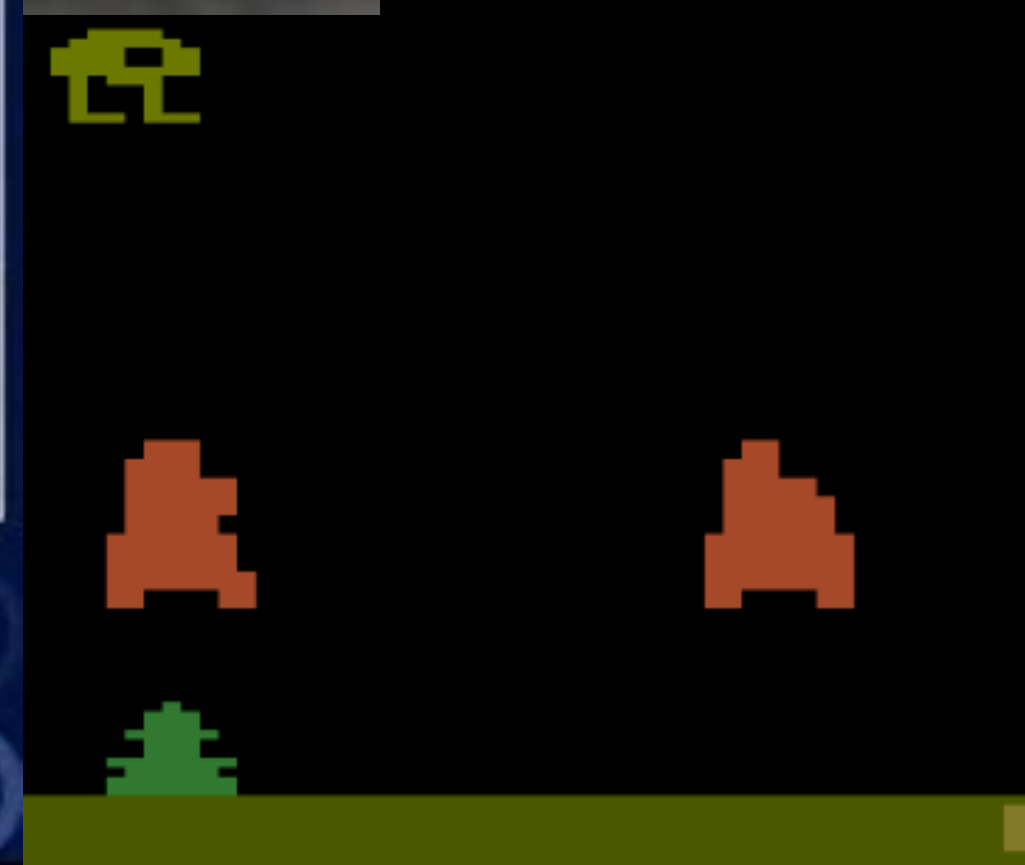
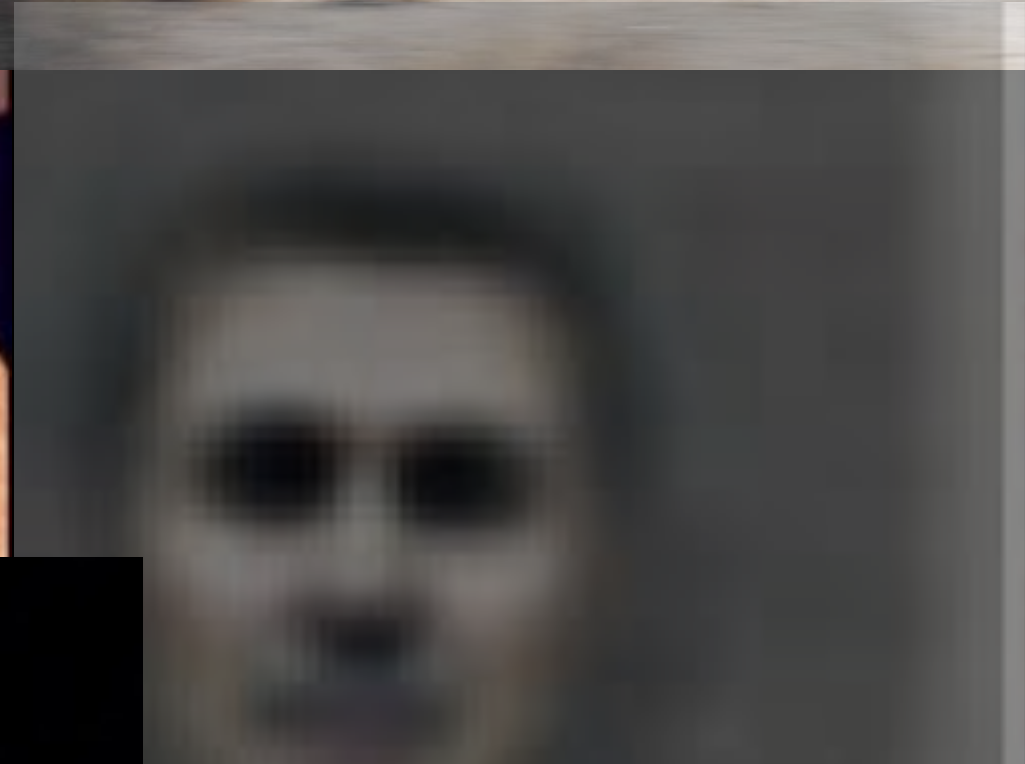


Balancing risks and opportunities: future directions and responsible innovation

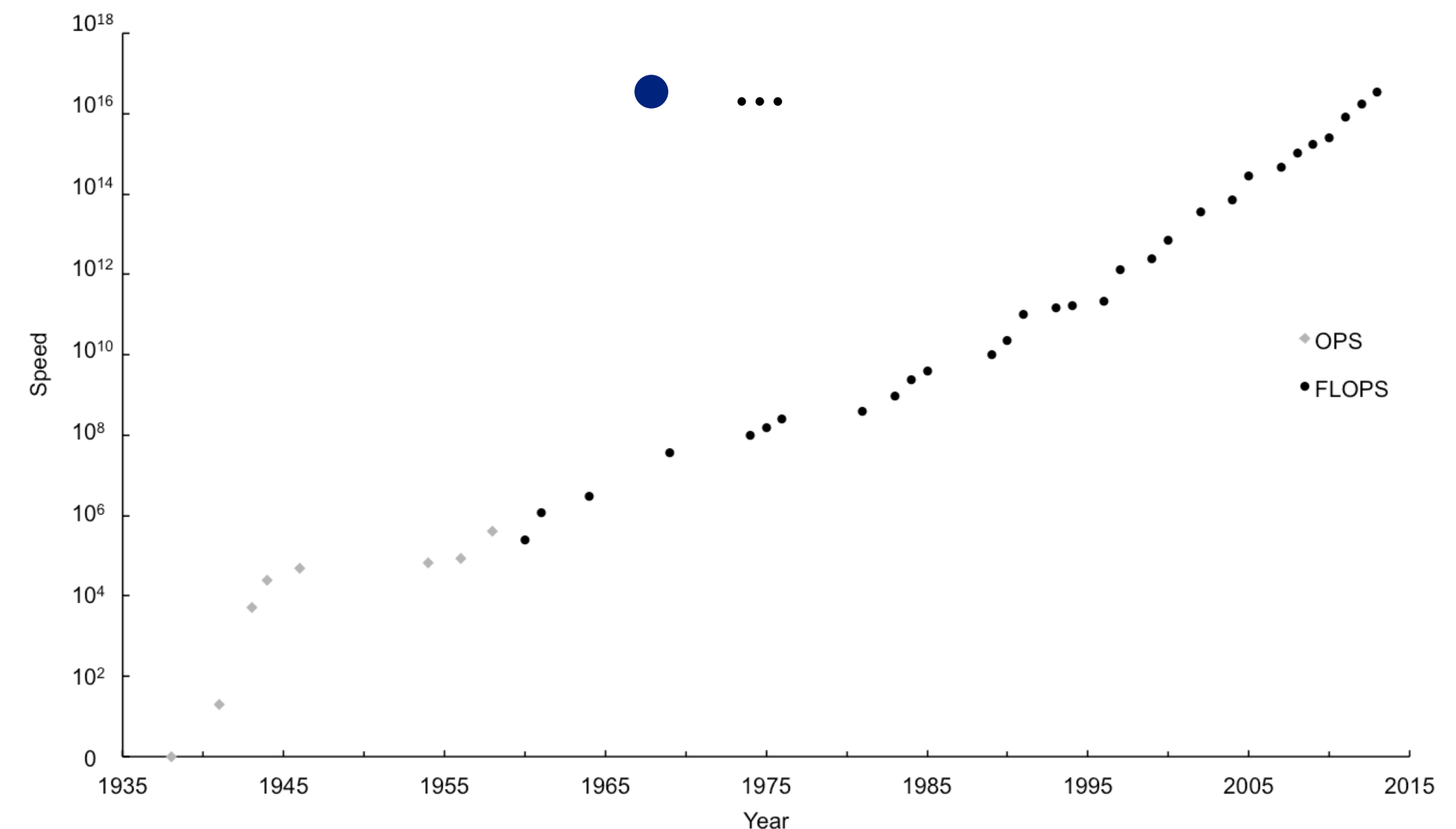
Dr. Niel Bowerman

Assistant Director, Future of Humanity Institute
University of Oxford

Thanks to Prof. Nick Bostrom



- Decision theory
- First-order logic
- Heuristic search
- Decision trees
- Alpha-Beta Pruning
- Hidden Markov Models
- Policy iteration
- Backprop algorithm
- Evolutionary algorithms
- Support vector machines
- Hierarchical planning
- Algorithmic complexity theory
- TD learning
- Bayesian networks
- Big Data
- Variational autoencoders
- Convolutional neural networks
- Deep learning
- Neural Turing machines
- ...



Science



SPECIAL ISSUE

ARTIFICIAL
INTELLIGENCE

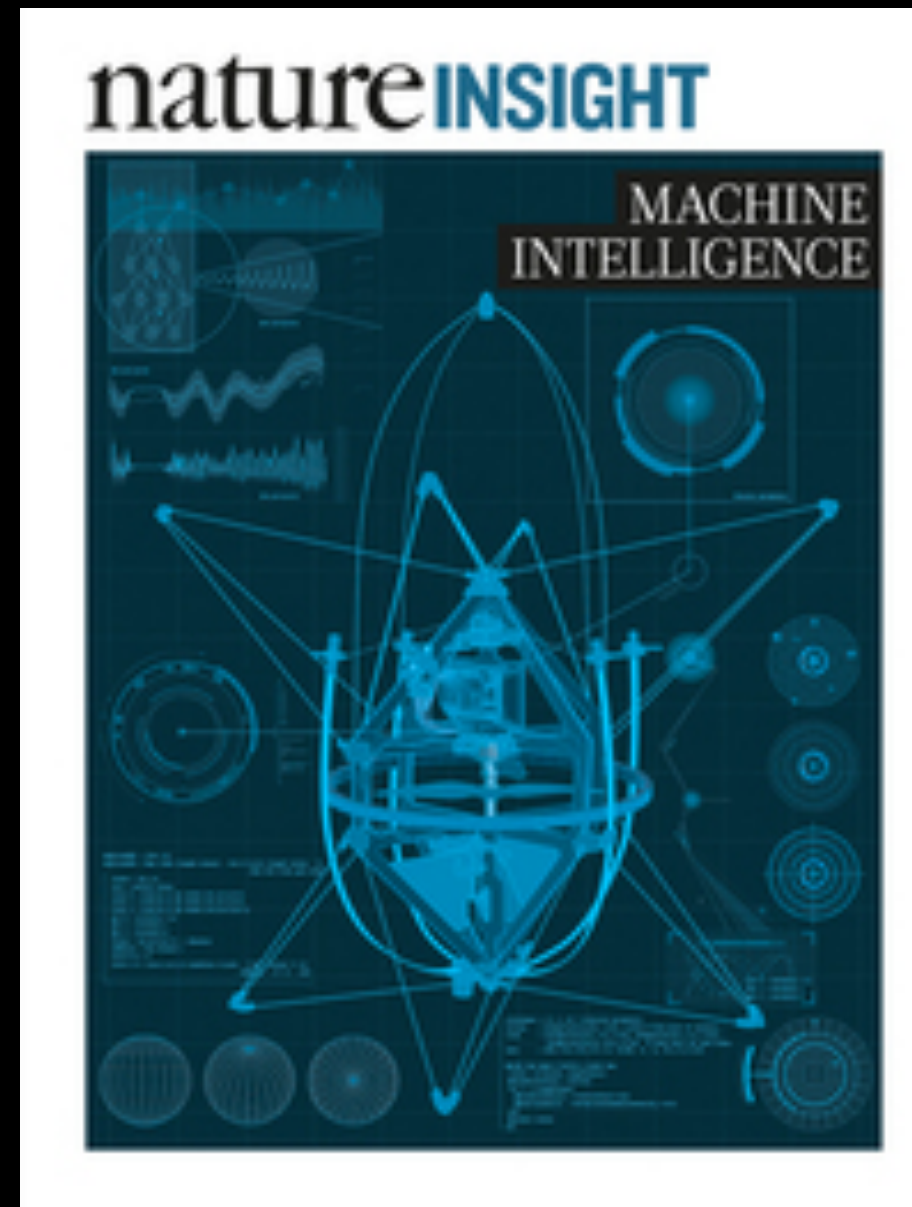
What it takes to end an
AIDS epidemic p. 226

Polar bears suffer through
lean summers p. 255

Sperm pro
of mutant

17 JULY 2015
sciencemag.org

AAAS





Google

facebook

 Microsoft



AI policy areas

- Data & privacy
- Autonomous systems & liability
- Automation & unemployment
- Military, security, and geopolitical coordination

AI policy areas

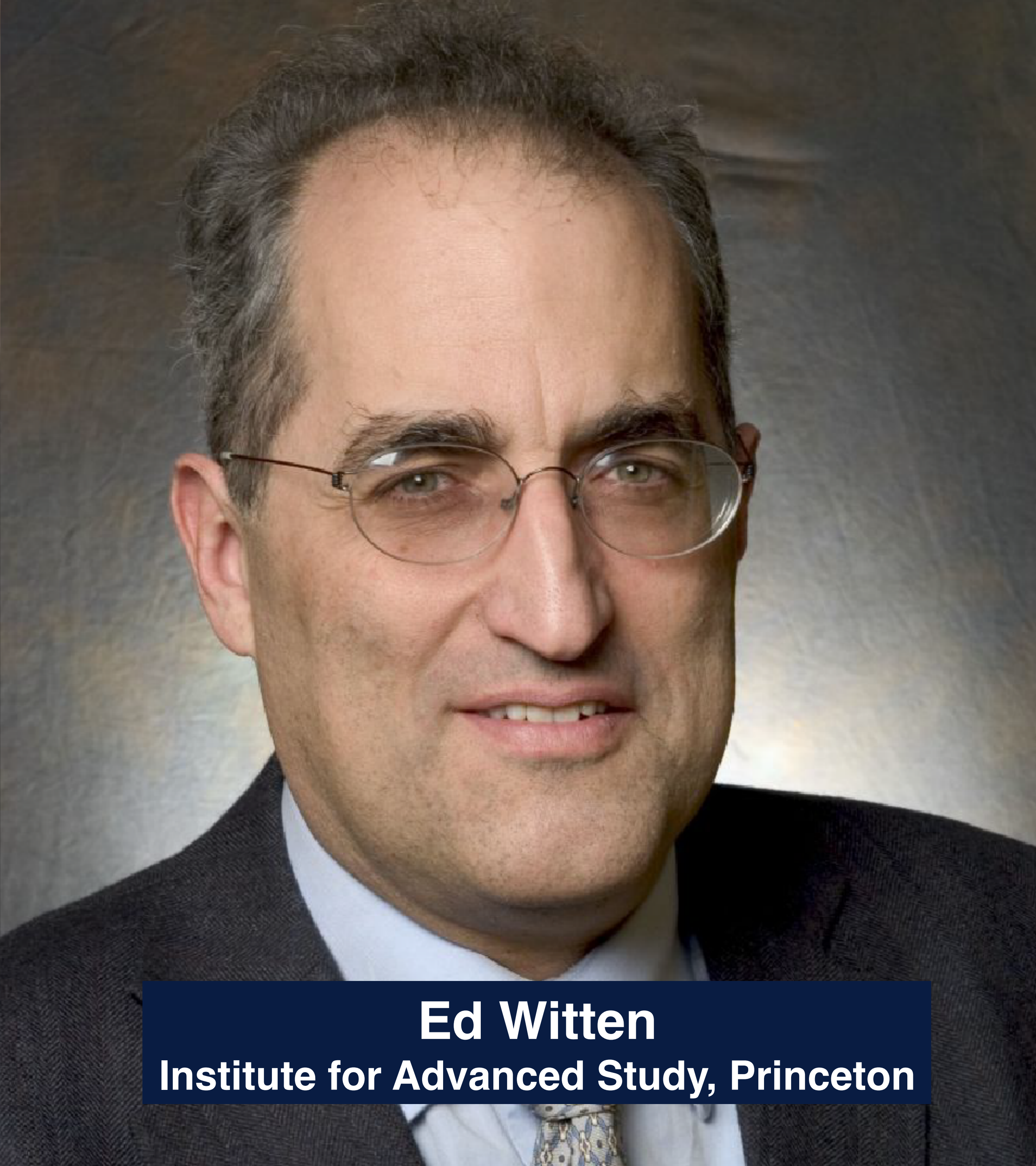
- Data & privacy
- Autonomous systems & liability
- Automation & unemployment
- Military, security, and geopolitical coordination

But what happens if we actually succeed?



Kanzi

Great Ape Trust, Des Moines, Iowa



Ed Witten

Institute for Advanced Study, Princeton



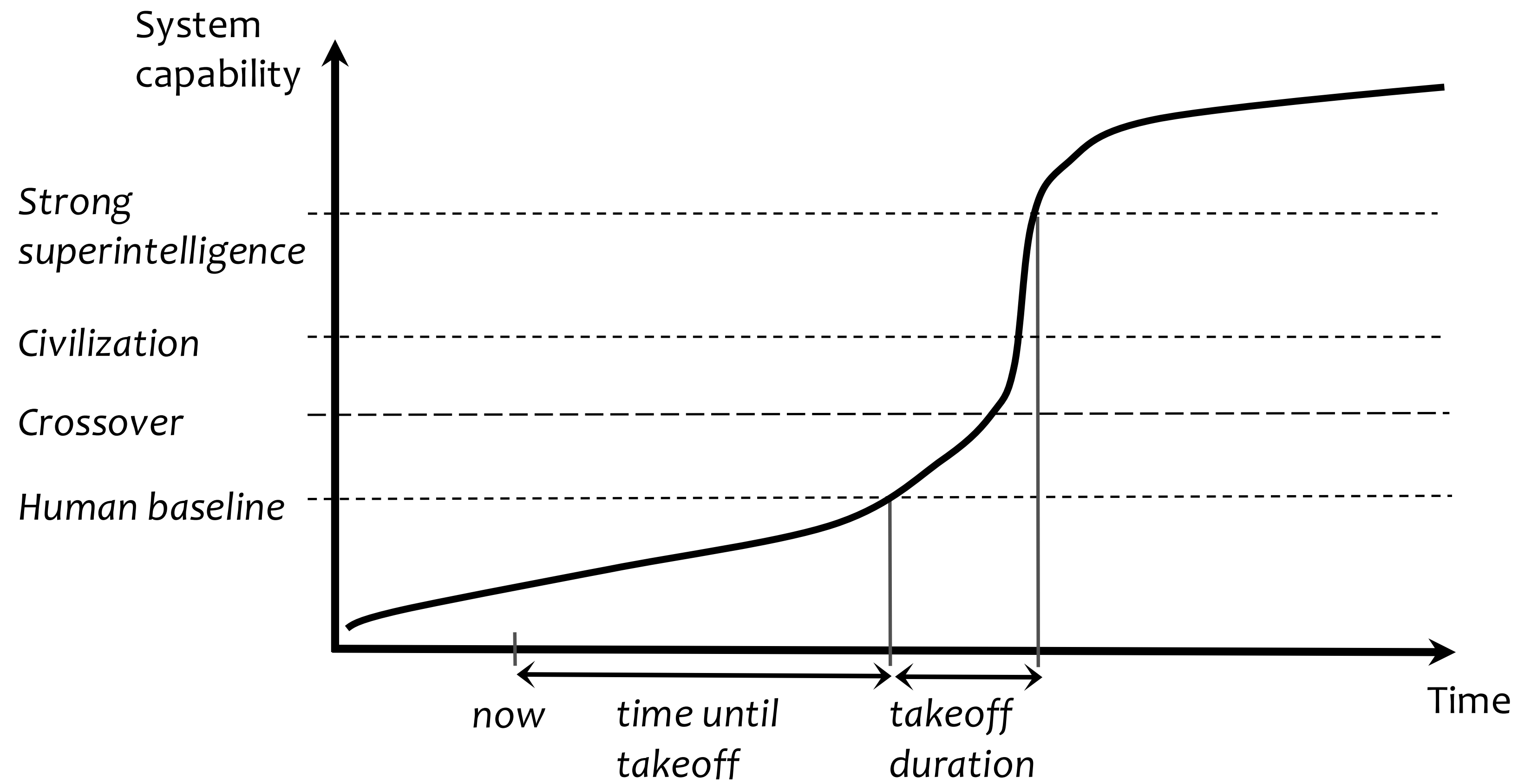
Pan Troglodytes



Homo Sapiens

When will HLMI be achieved?

	10%	50%	90%
TOP100	2024	2050	2070
Combined	2022	2040	2075



**Village
idiot**



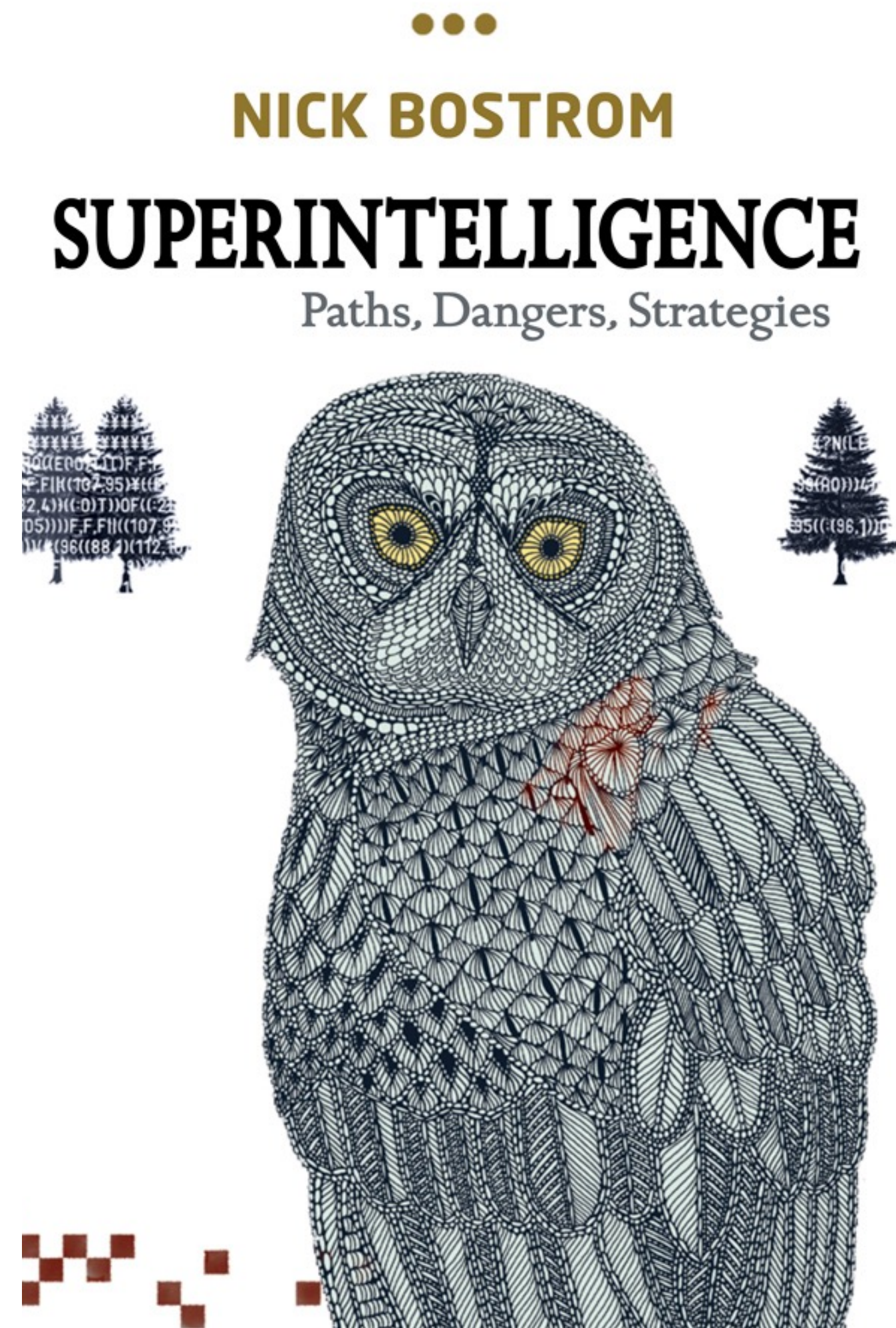
**Ed
Witten**





The control problem

- Superintelligence raises unique challenges
- Plausible scenarios in which superintelligent systems eventually become very powerful
- And apparently adequate control mechanisms fail



Recommendations

- Be careful not to take actions which are based on an implicit premise about AI having a clear upper limit in future capability.
- Groups investigating these issues may benefit from focusing on broad principles and not narrow recommendations. These should be flexible in application to new developments and increased capability.
- EU should keep a broadly global view in its approach on AI. It is possible that AI advances will be widely destabilizing, and that some nations will not adapt well. Both the EU, and the rest of the world, will benefit if the EU can lay out broad, cooperative, and replicable frameworks which favour stability, economic well-being, and coordination.