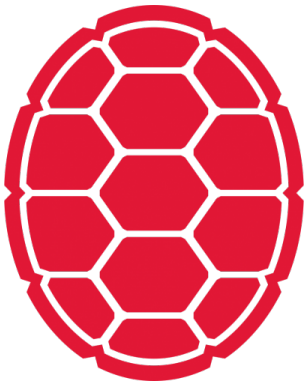


Will AI Kill Us All?

Machine Learning: Jordan Boyd-Graber
University of Maryland
SCENARIOS

Ideas Adapted from Rodney Brooks

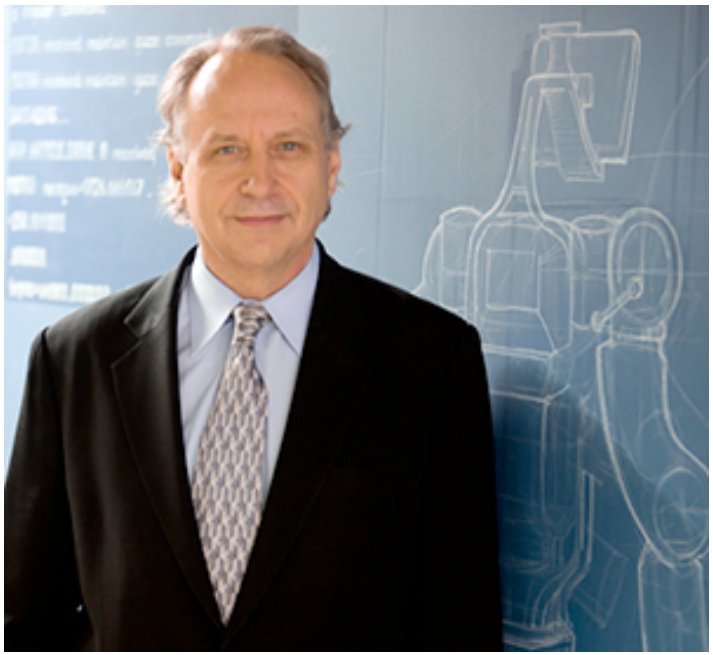
What is the threat?

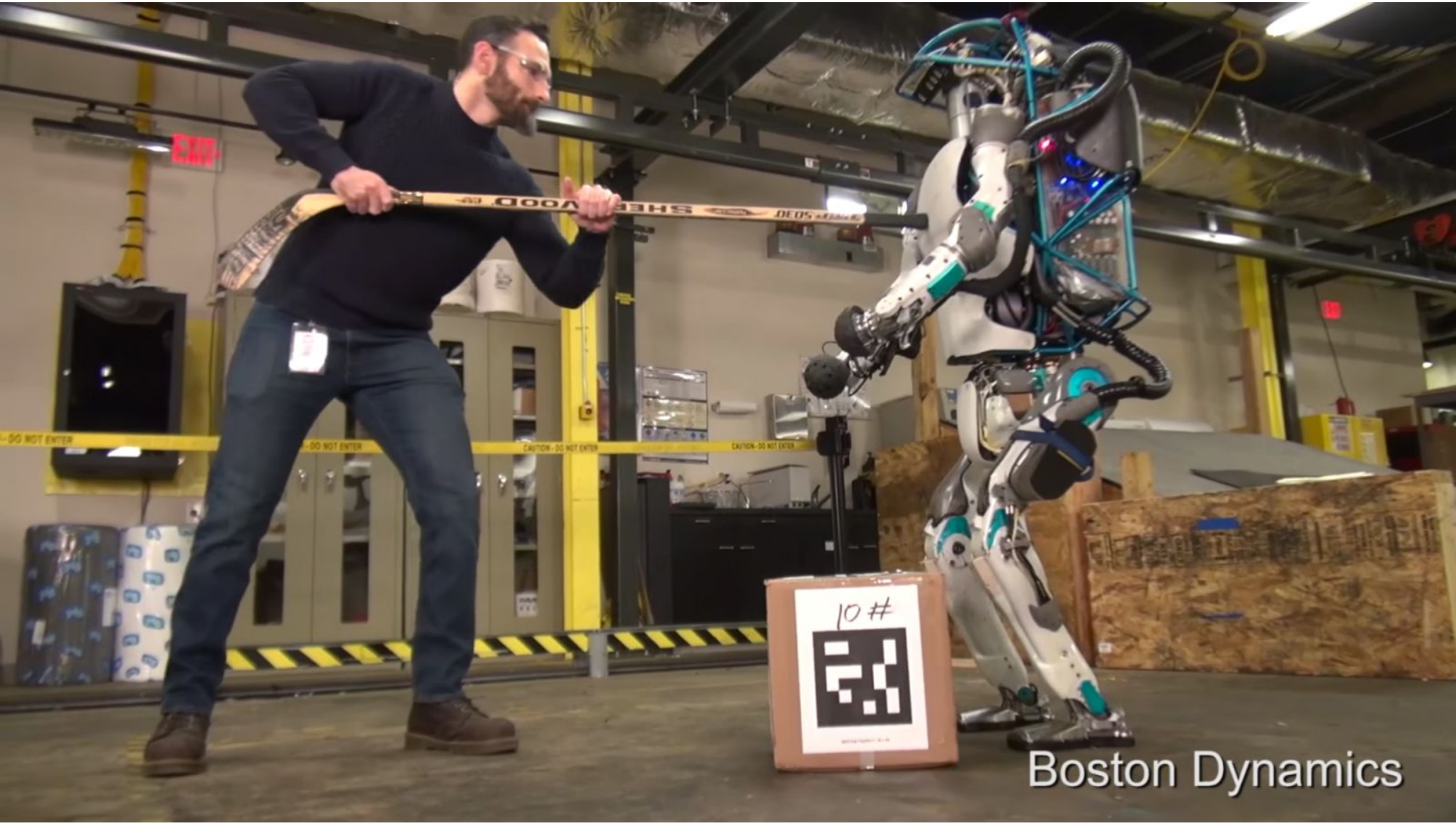
- General Artificial Intelligence (AGI)
- Singularity
- Misaligned Objectives
- Pure Evil
- Societal Upheval

It will take a while . . .

Rodney Brooks

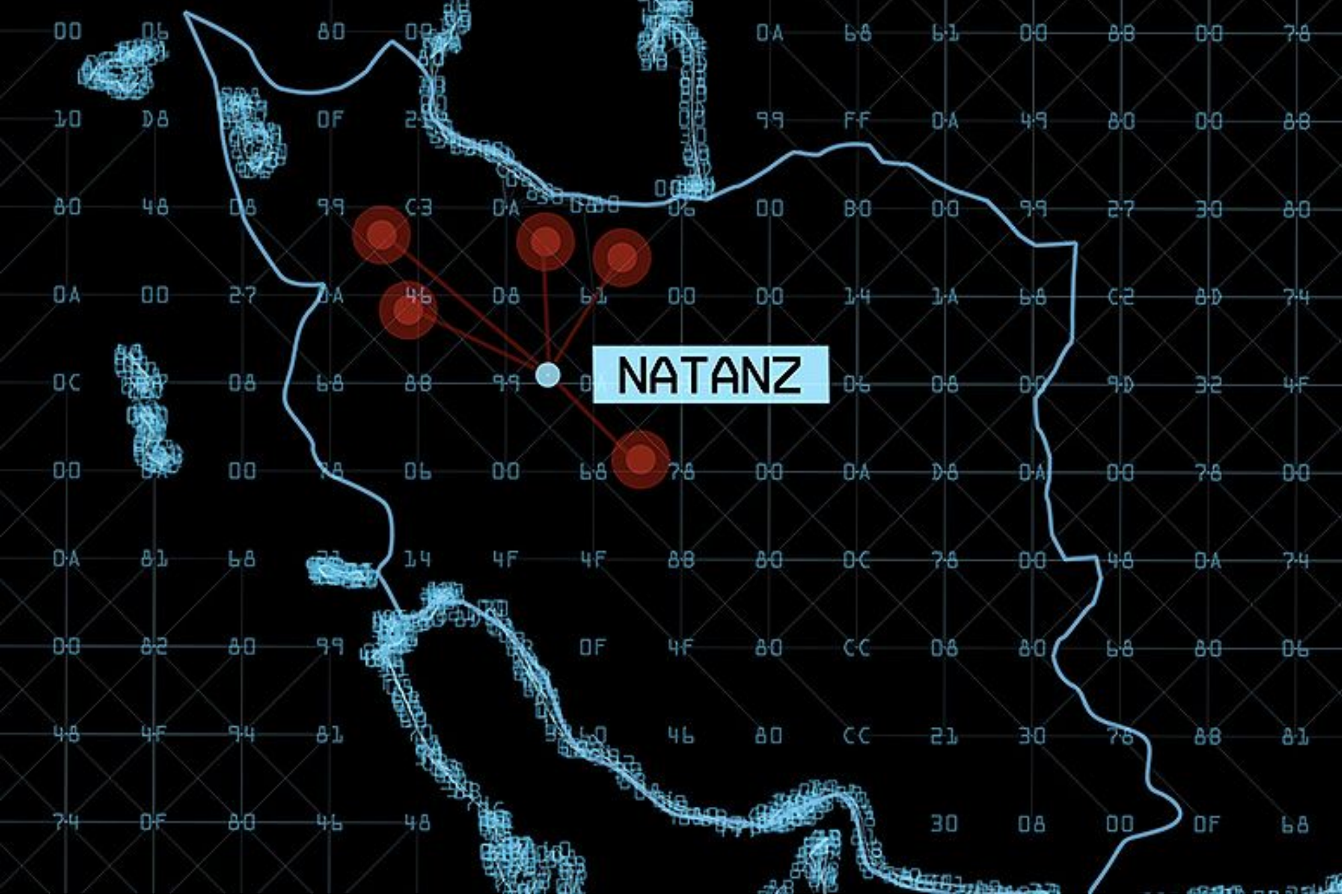
Most technologies are overestimated in the short term. They are the shiny new thing.
Not all technologies get underestimated in the long term, but that is most likely the case for AI. The question is how long is the long term.



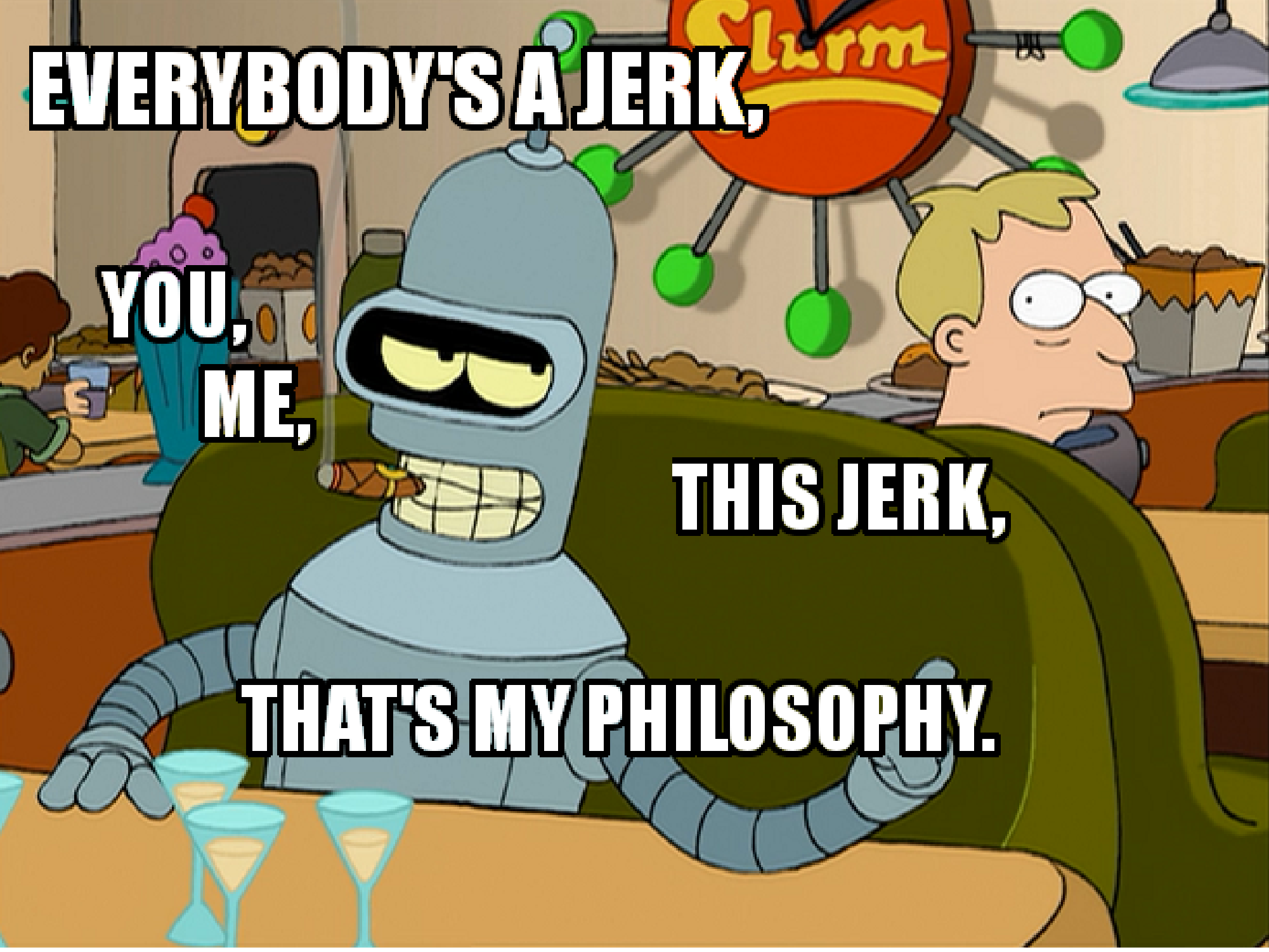


Humans are jerks

Humans are jerks

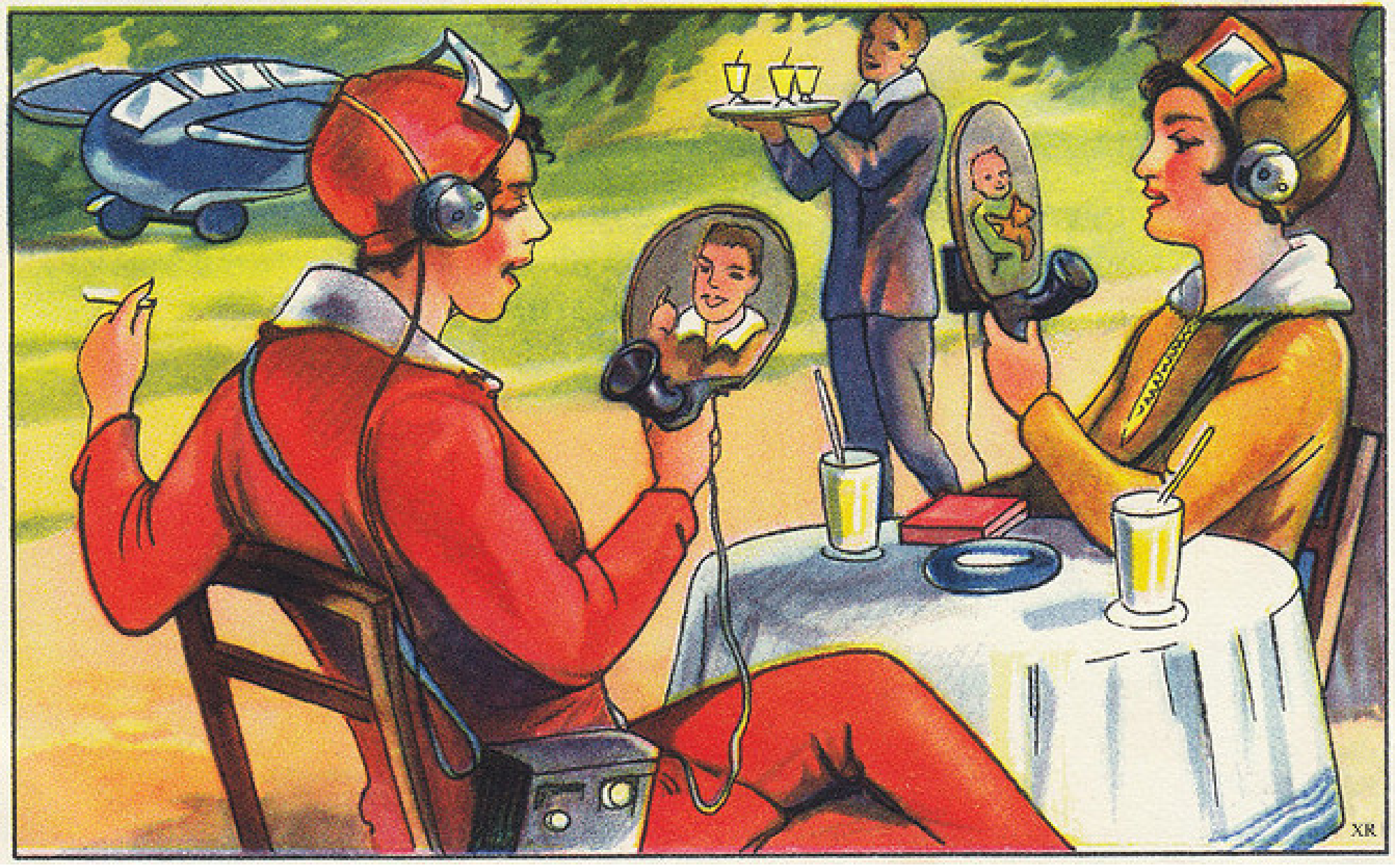


Governments are bigger jerks

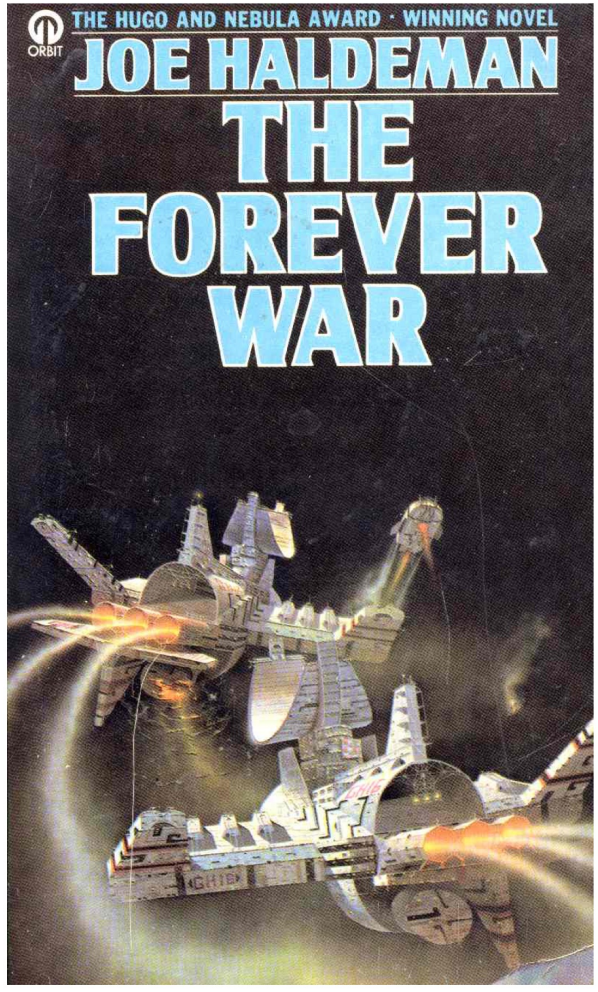

EVERYBODY'S A JERK,
YOU,
ME,
THIS JERK,
THAT'S MY PHILOSOPHY.

Prediction 1: Building Immune System

- Bad things will happen
- AI may be involved
- But prime mover will be human
- Will cause us to take cybersecurity seriously



Technology changes but society doesn’t

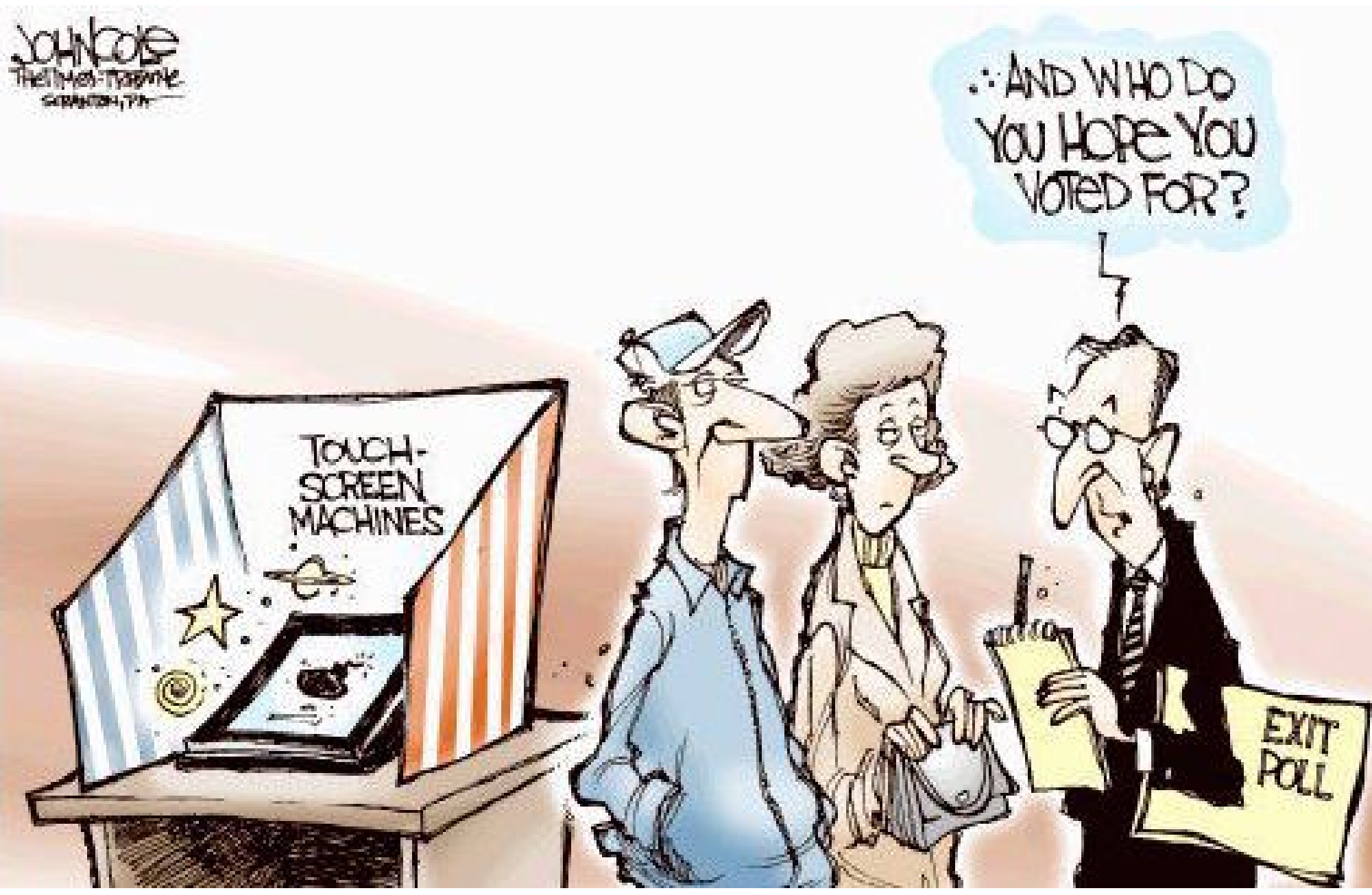


A book that gets it right

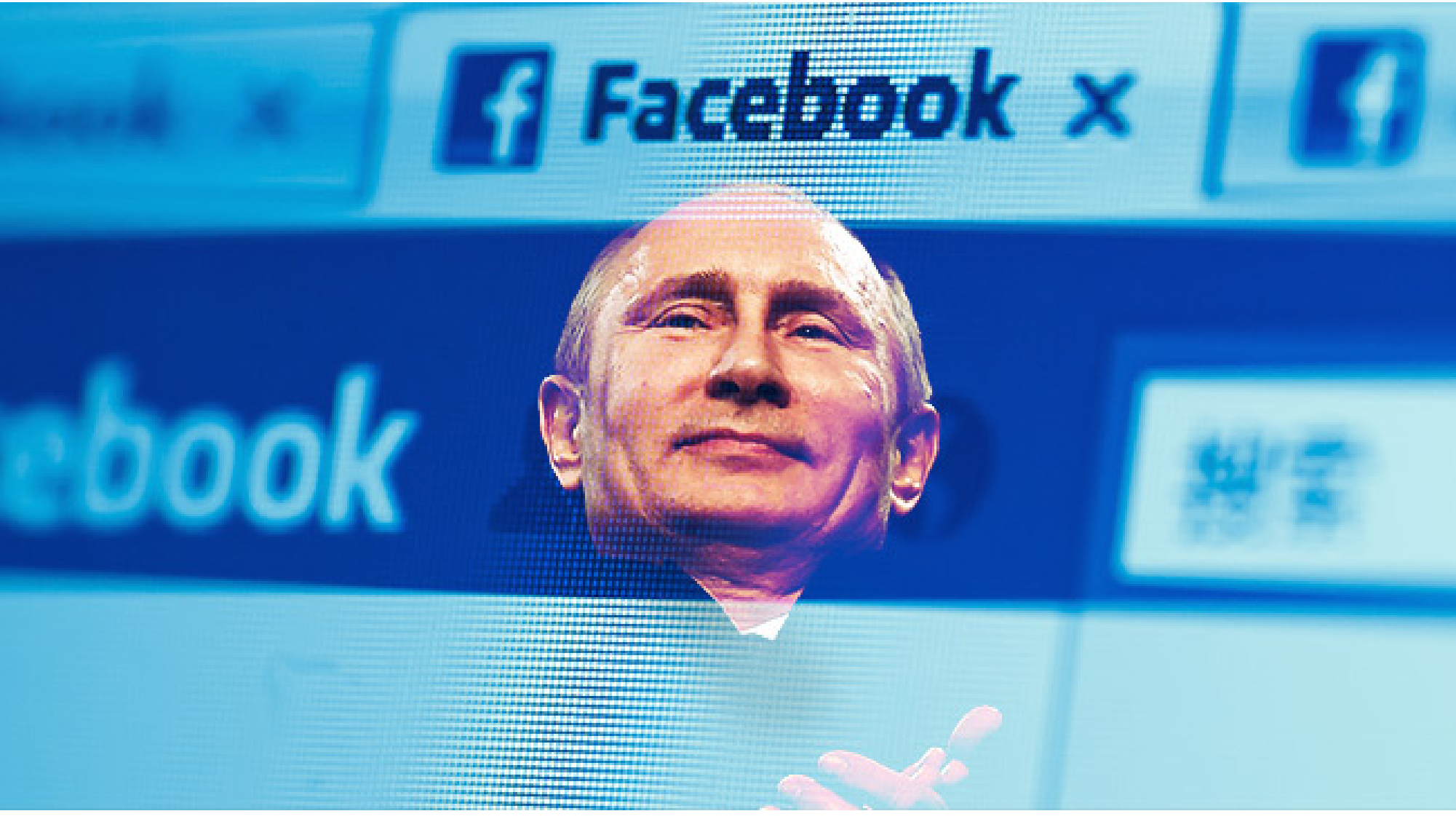
Prediction 2a: Physical Harm from Special-Purpose AI

- Special-purpose AI will be used for terror
- Will be scary
- We may overreact
- But utility clear
- Not an exestential threat (c.f. nuclear weapons)

But that isn’t a new threat . . .



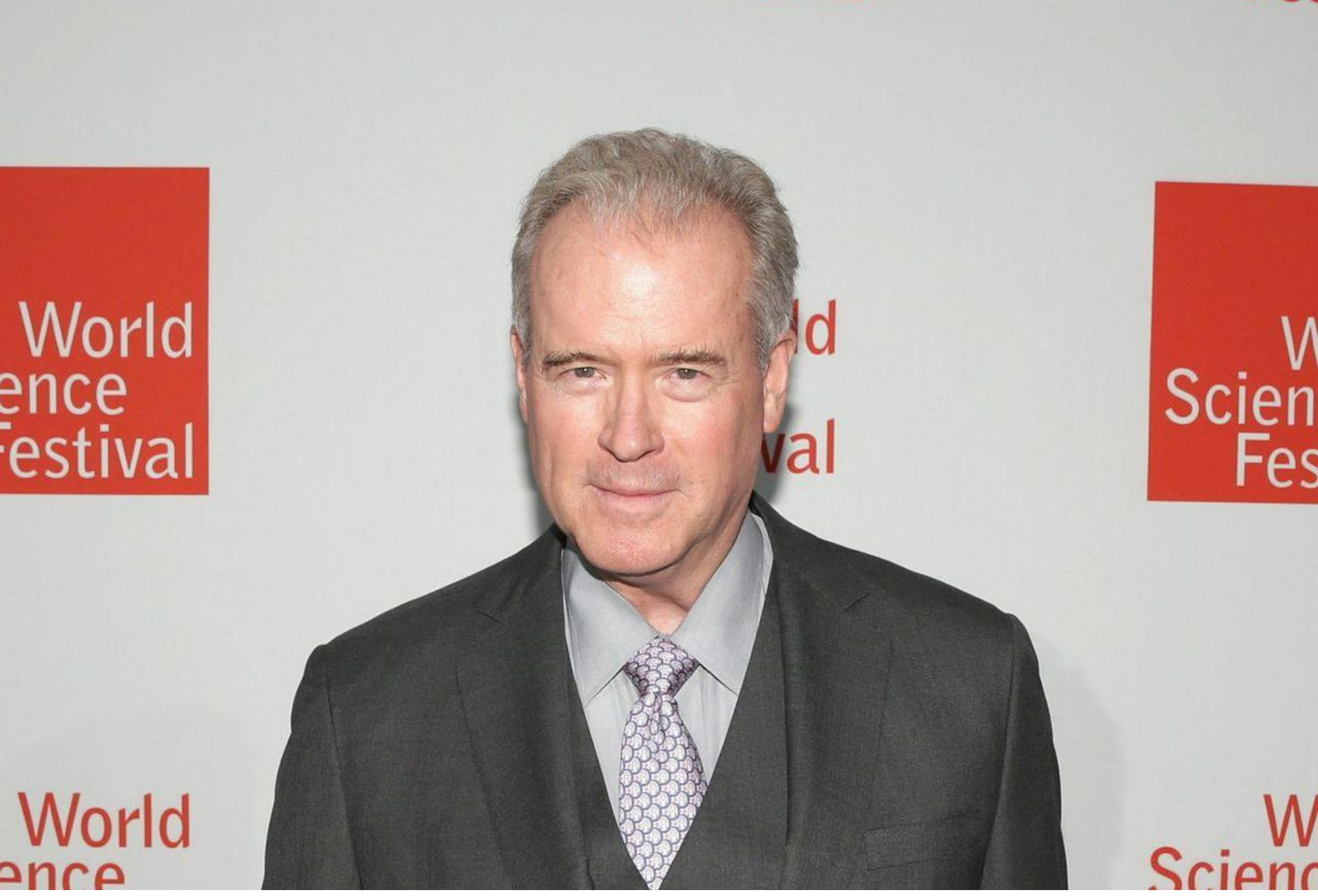
Political process at risk internally



Political process at risk externally

Prediction 2b: Social Manipulation

- Will happen (is happening?)
- But will be slow
- Will be scaled up version of what humans already do
- Anti-singularity: Humans are slow
- Hopefully we will build defenses



Renaissance, Two Sigma, Voleon

Prediction 2c: Economic Catastrophe

- Fear: paperclip AI
- Is already happening
- But happening from many directions
- Efficient markets (**good thing**!)
- Human irrationality

What about AI that does everything?

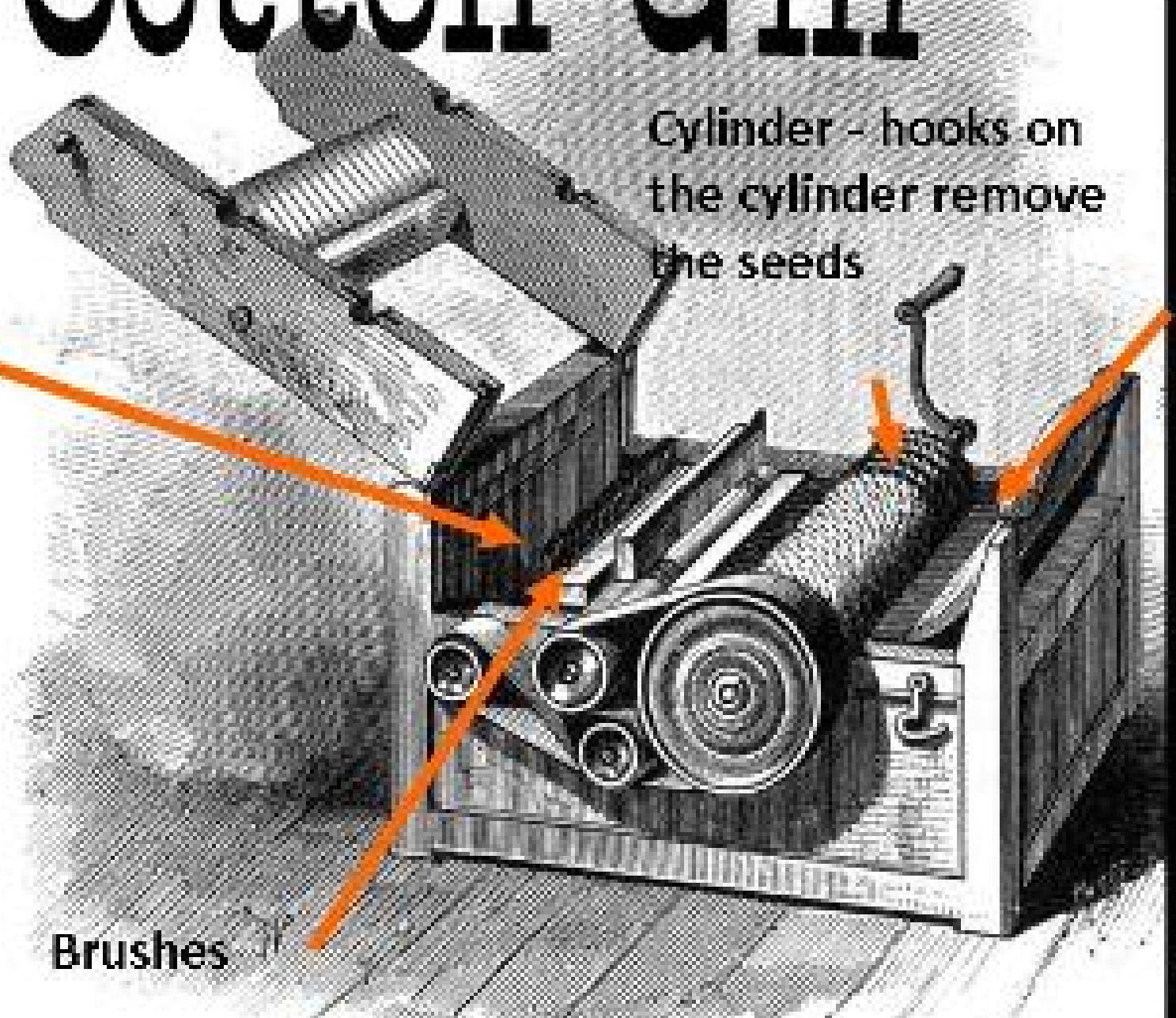
Prediction 3: AGI

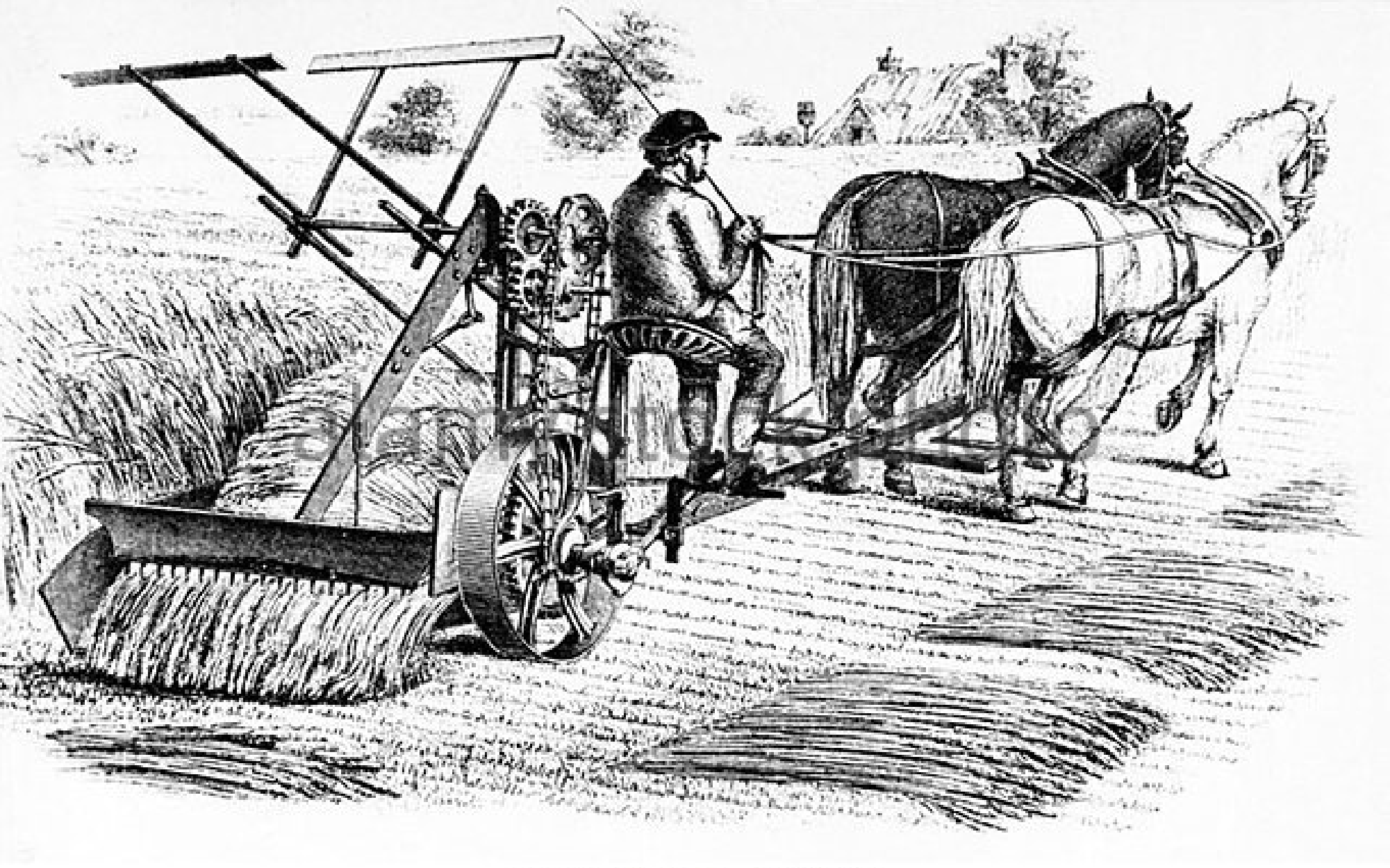
- If AGI happens, society will be very different
- Special purpose AI will exist to prevent takeover (previous prediction)
- Will have seen smaller, more specific challenges
- Fearmongering ignores this long-term context (which is hard to predict)
- Analogy: immune system

Prediction 3: AGI

- If AGI happens, society will be very different
- Special purpose AI will exist to prevent takeover (previous prediction)
- Will have seen smaller, more specific challenges
- Fearmongering ignores this long-term context (which is hard to predict)
- Analogy: immune system
- So, nothing to worry about?

Cotton Gin

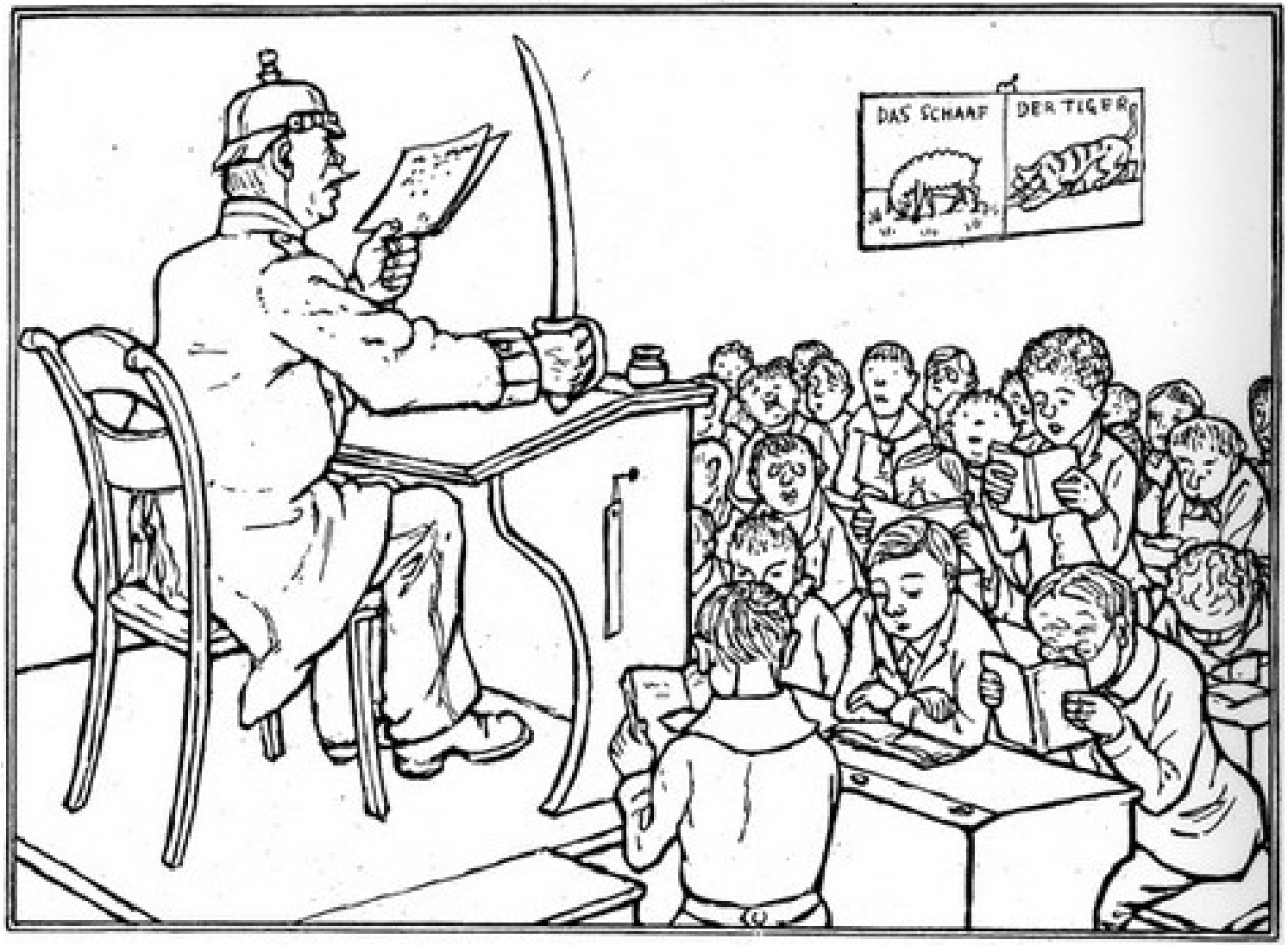

Cylinder - hooks on the cylinder remove the seeds
Clean cotton here.
Cotton bolls with seeds are placed here.
Brushes



Don’t fear the reaper

bp k bildarchiv preussischer kulturbesitz

Fortschritt im Schulwesen

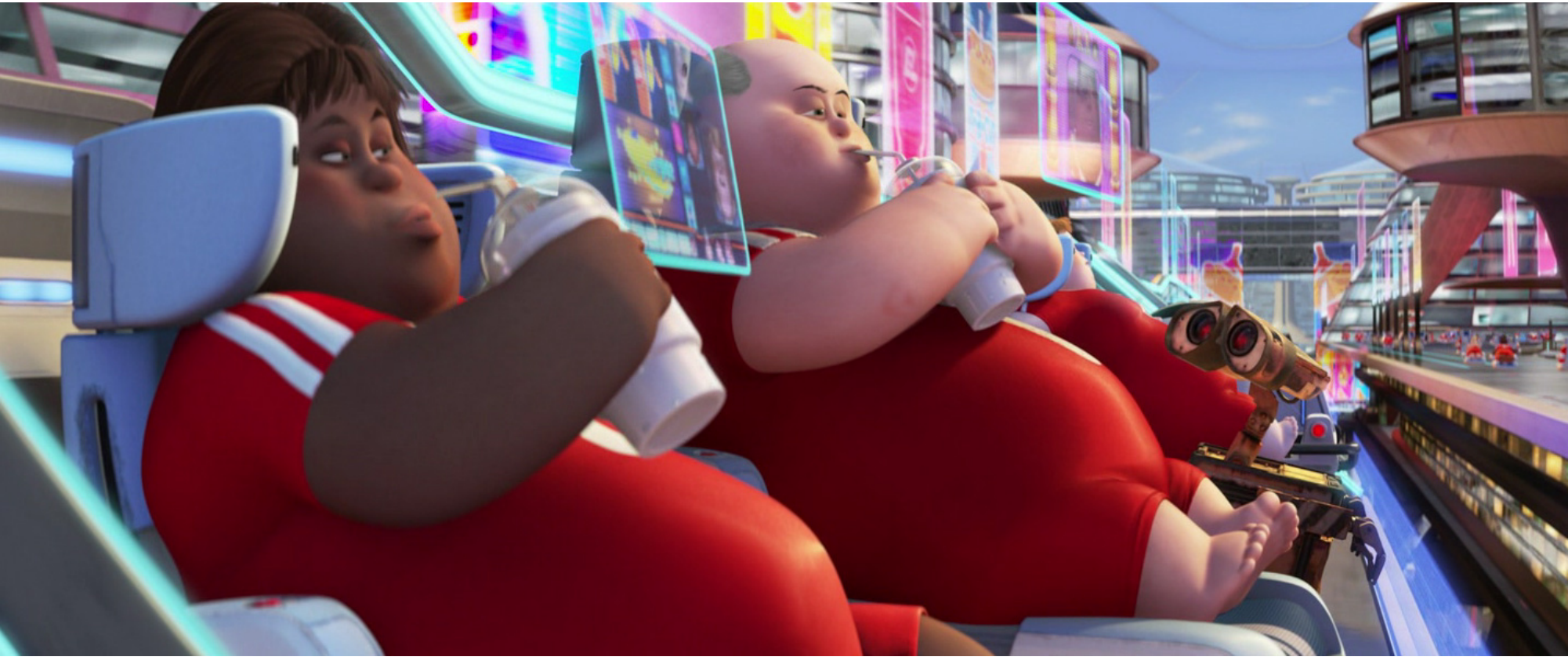


An den preußischen Schulen soll jetzt als neues Unterrichtsfach Staatsbürgerkunde eingeführt werden. Es ist wohl selbstverständlich, daß der Unterricht von Schutzleuten erteilt wird.

Social reforms paved the way

Industrialization reordered society

Automation **slowly** taking hold



Are we prepared for social changes?

Prediction 4

- Education has not changed from what we needed for assembly line
- AI could worsen inequality
- Society needs to restructure
  - Bad: Massive unemployment and misery
  - Good: Redistribution
  - Great: Allow humans to focus on creativity

(-11.177339) a man is holding a skateboard in the air

ML sucks at creativity

What should ML researchers do?

- Support creativity
  - Support arts
  - Support education that makes careers less automatable
  - Support more personalized experiences
  - Create ML that enhances humans' inante abilities

What should ML researchers do?

- Support creativity
  - Support arts
  - Support education that makes careers less automatable
  - Support more personalized experiences
  - Create ML that enhances humans' inante abilities
- Help others understand how much ML sucks
- Build the foundations of the world your (grand)children want to live in
- Make AI interpretable and accountable