THREAT MODELING GLOBAL CATASTROPHIC RISKS

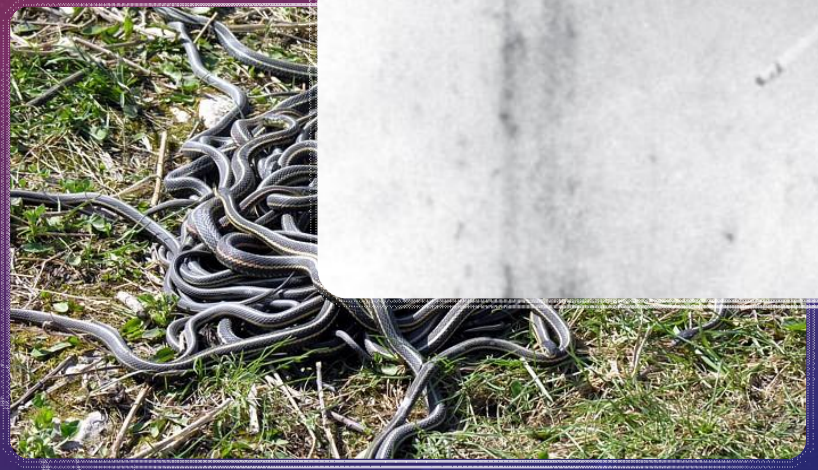
LUKE DONOHO, M.S

OH MY GOD WE'RE ALL GOING TO F*!&@#^$%ING DIE

LUKE DONOHO, M.S.

AGENDA

- Existential Risk
- Threat Modeling
- Artificial Intelligence
- Nanotechnology
- Threat Modeling Existential Risk




Garuda Indonesia

EXISTENTIAL RISK

Future of Humanity Institute

OXFORD MARTIN SCHOOL

UNIVERSITY OF OXFORD

Global Challenges Foundation

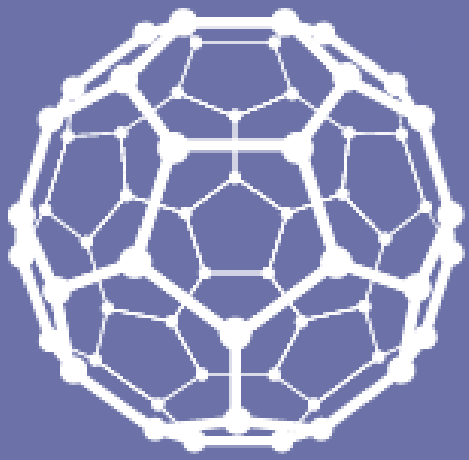
Global Challenges Foundation

GLOBAL CATASTROPHIC RISKS

- Reliability of AI Predictions
- A single dominant AI or many
- How intelligent AI will become
- Whether extremely intelligent AI can be controlled
- Whether whole brain emulations will come before AI

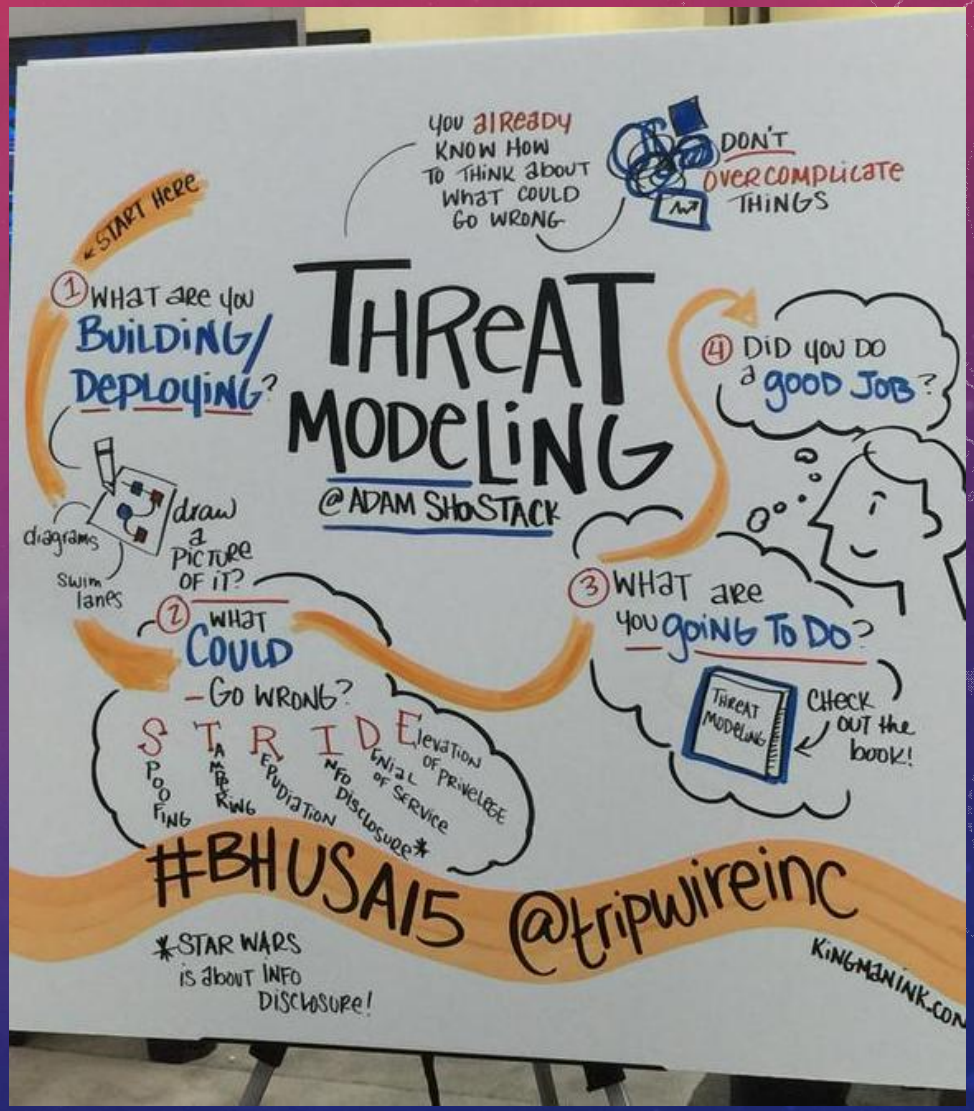
Artificial Intelligence

- Timeline for development
- Order of research progression
- How quickly weapons can be assembled
- Surveillance capabilities of tools
- Whether tools are made outside of human control

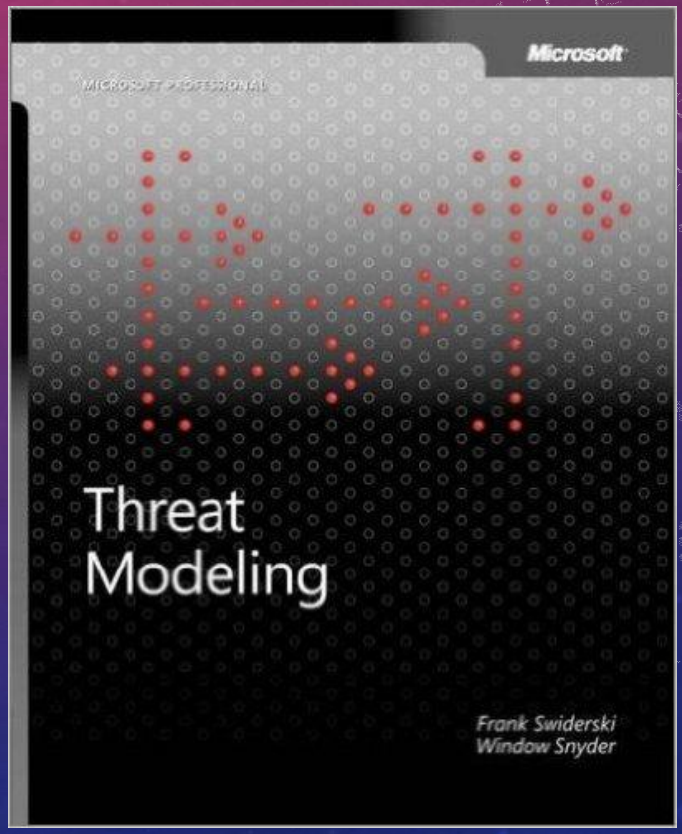


Nanotechnology

THREAT MODELING



THREAT MODELING – WINDOW SNYDER



“Threat modeling is a methodology for identifying security vulnerabilities, for identifying the risks of a security vulnerability within that application”

Window Snyder

THREAT MODELING – WINDOW SNYDER

THREAT MODELING – WINDOW SNYDER

“Here's a pattern, and if we implement one architectural change we can eliminate all these vulnerabilities.”

Window Snyder

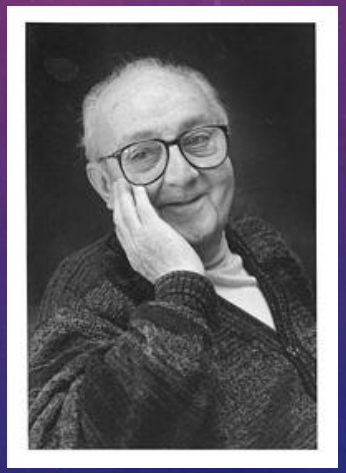
The goal, she said, is to remove whole categories of vulnerabilities.

THREAT MODELING – ADAM SHOSTACK

THREAT MODELING

https://www.youtube.com/watch?v=KLpgaoD8ySM

“All models are wrong, some models are useful”



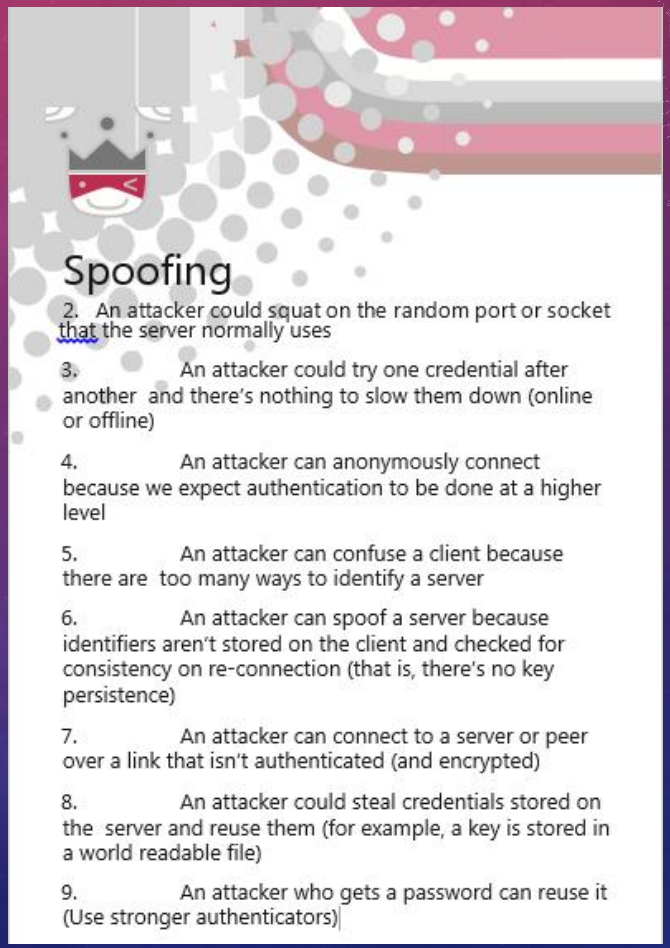
- George Box

THREAT MODELING - STRIDE

- Spoofing
- Tampering
- Repudiation
- Information Disclosure
- Denial of Service
- Elevation of Privilege

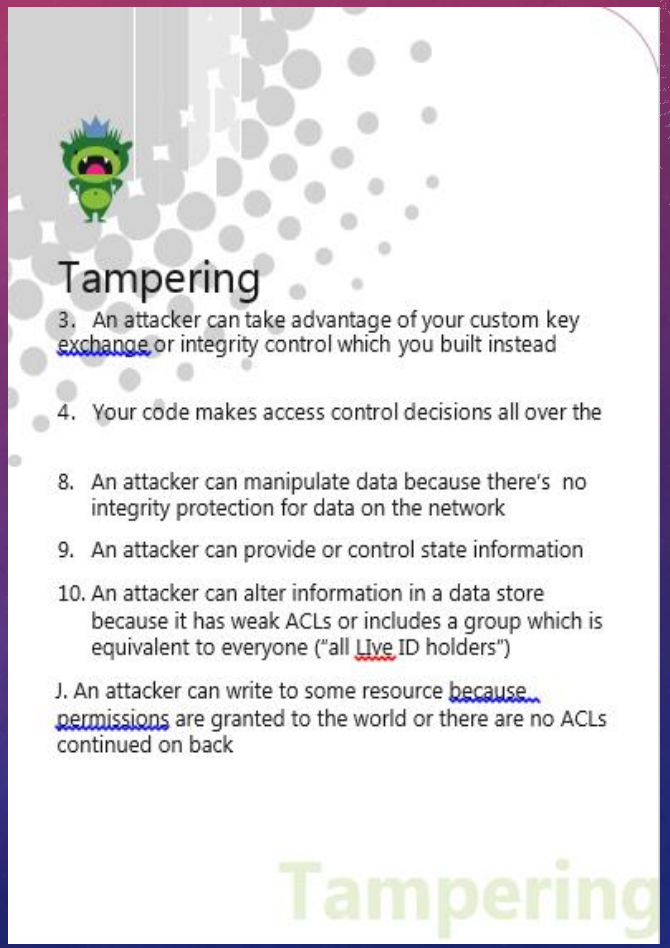
THREAT MODELING – STRIDE SPOOFING

Impersonating something or someone else



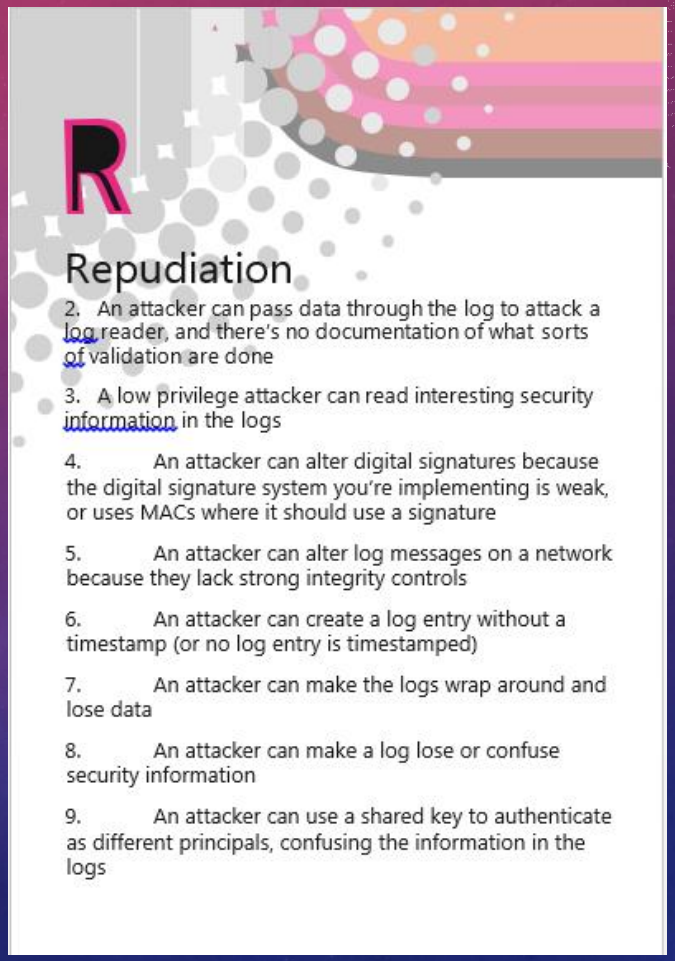
THREAT MODELING – STRIDE TAMPERING

Modifying Data or Code



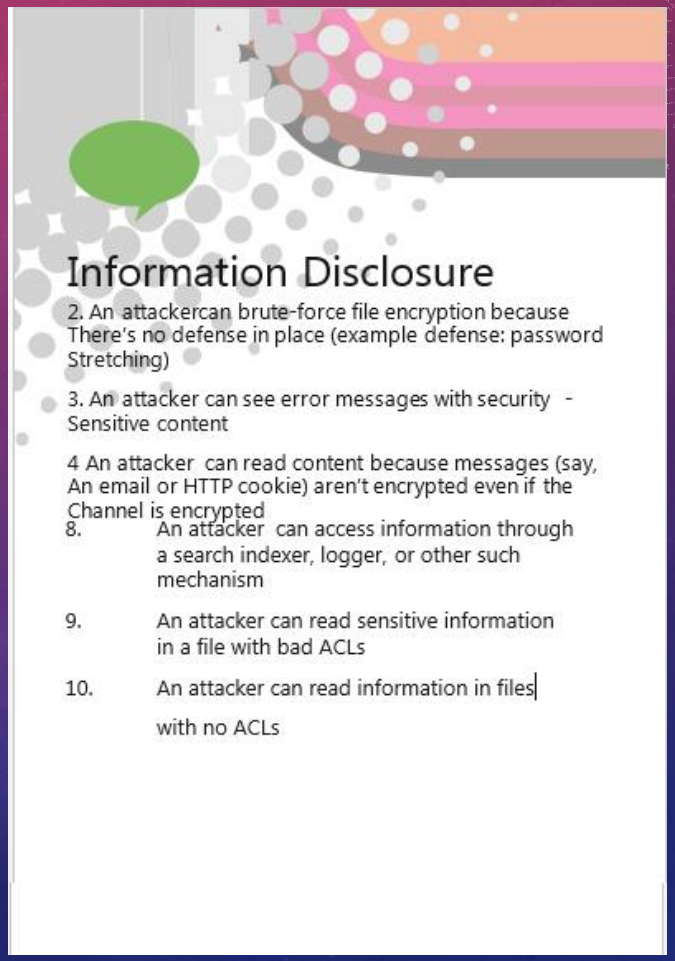
THREAT MODELING – STRIDE REPUDIATION

Claiming to have not performed an action



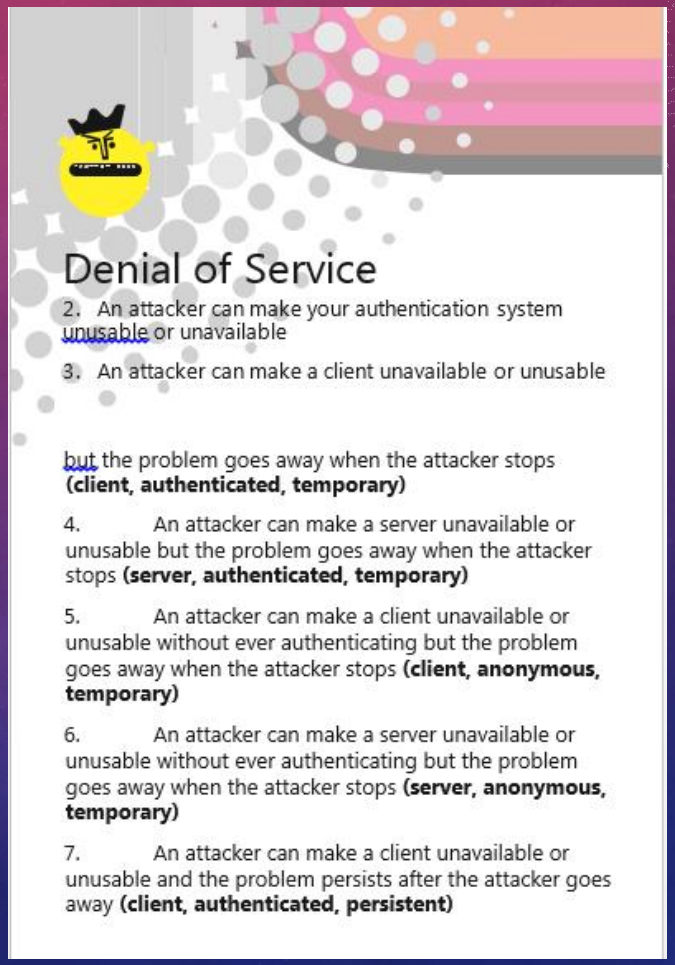
THREAT MODELING – STRIDE INFORMATION DISCLOSURE

Exposing information to someone who should not have access



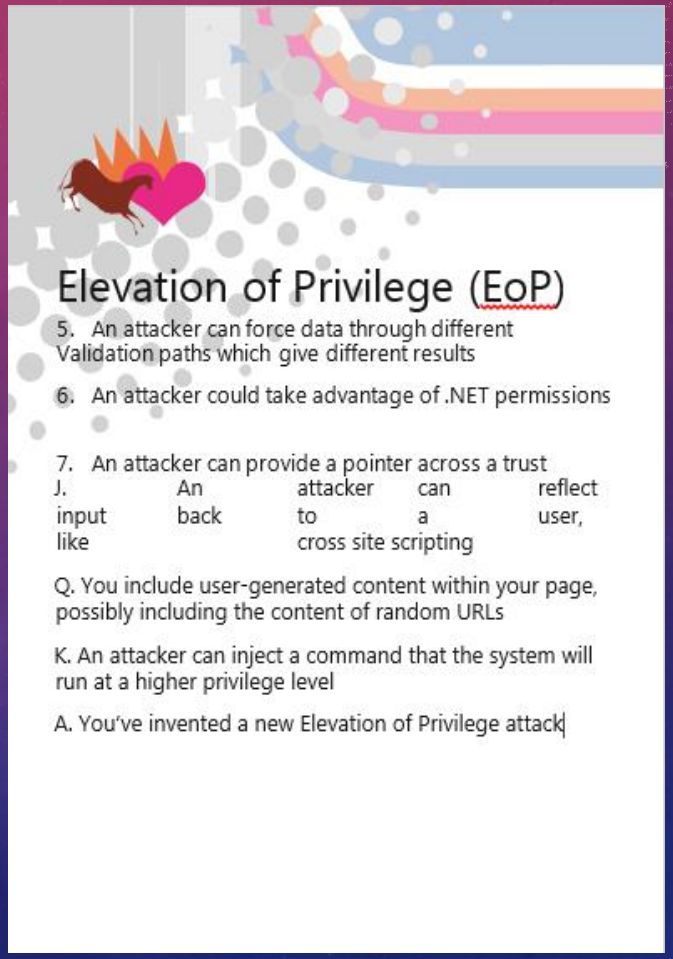
THREAT MODELING – STRIDE
DENIAL OF SERVICE

Deny or Degrade Services to Users

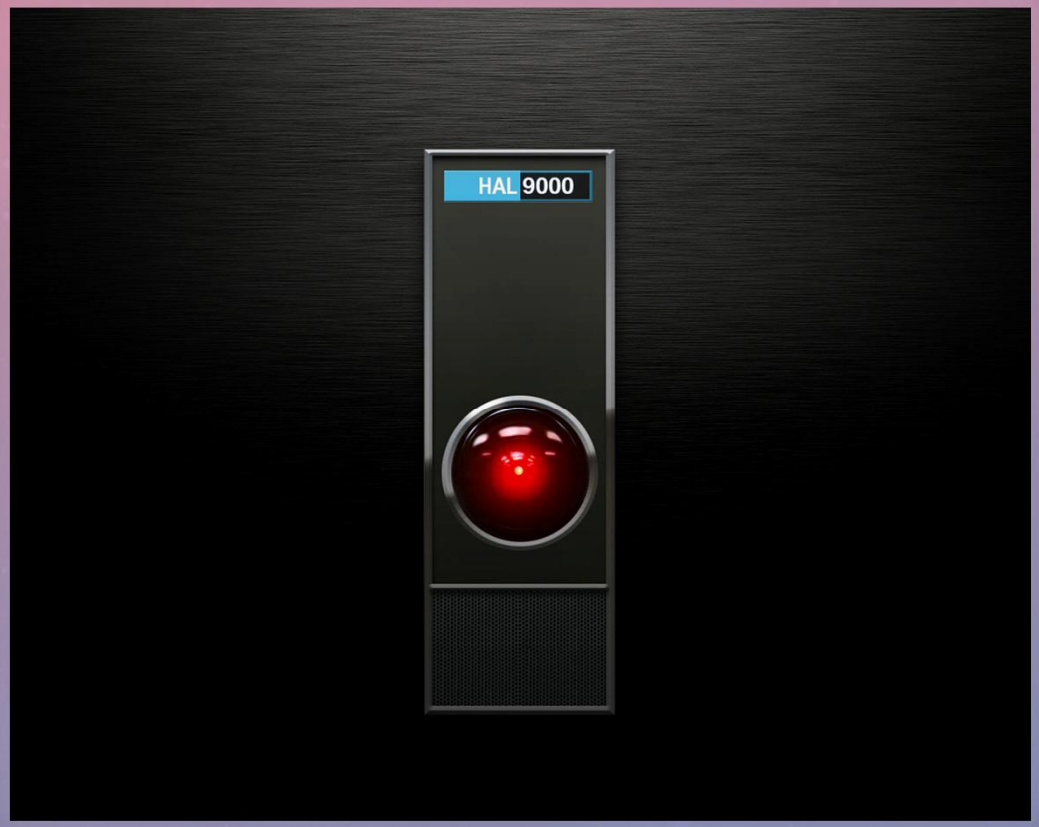


THREAT MODELING – STRIDE ELEVATION OF PRIVILEGE

Gain capabilities without proper authorization



ARTIFICIAL INTELLIGENCE

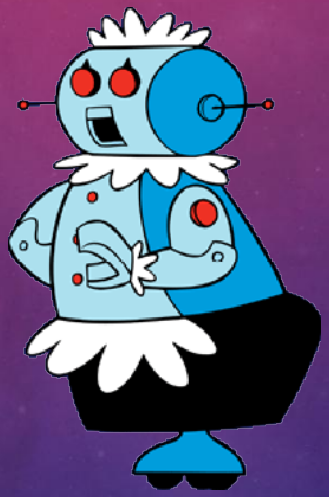


ARTIFICIAL INTELLIGENCE

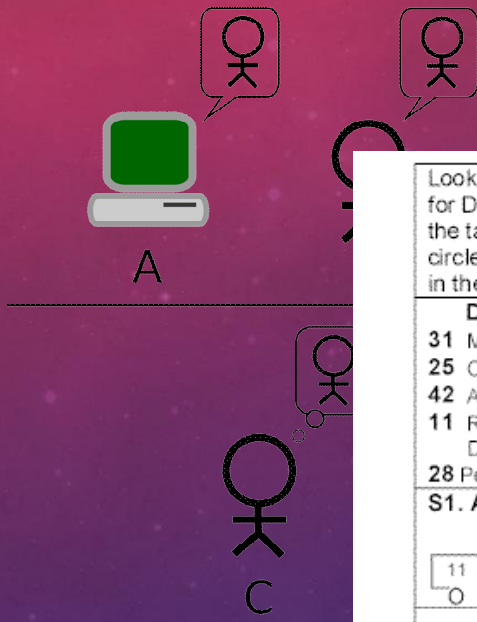
- Overview
- Ray Kurzweil
- Eliezer Yudkowsky
- Nick Bostrom

ARTIFICIAL INTELLIGENCE – WEAK AI

ARTIFICIAL INTELLIGENCE – STRONG AI (AGI)



ARTIFICIAL INTELLIGENCE – STRONG AI (AGI)



Look at the sample table of codes. The sample table lists codes for Department, Department Size, and Travel Budget. Below the table are two sample problems. In the problems, mark the circles below the codes that correspond to the information found in the table.

Department	Department Size	Travel Budget
31 Marketing	**1** 1-9 employees	**A** $500-999
25 Operations	**2** 10-19 employees	**B** $1,000-2,499
42 Accounting	**3** 20-39 employees	**C** $2,500-5,999
11 Research & Development	**4** 40-59 employees	**D** $6,000-7,999
28 Personnel	**5** 60 or more employees	**E** $8,000-10,000

S1. Accounting Dept., staff of 10, $3,000 Travel Budget.

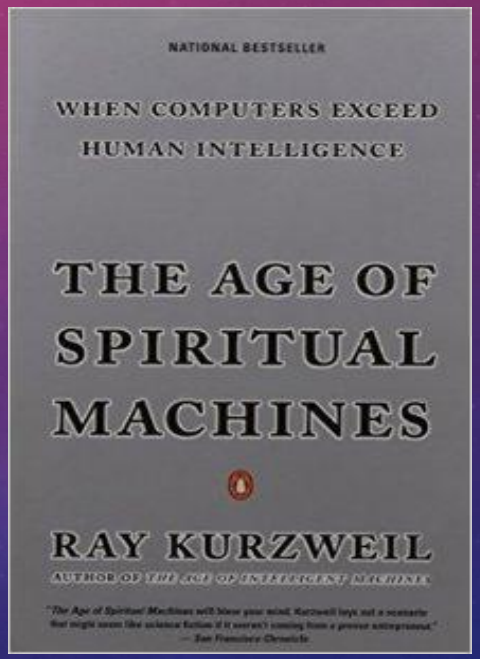
Department	Department Size	Travel Budget
11 25 28 31 42	1 2 3 4 5	A B C D E
○ ○ ○ ○ ●	○ ● ○ ○ ○	○ ○ ● ○ ○

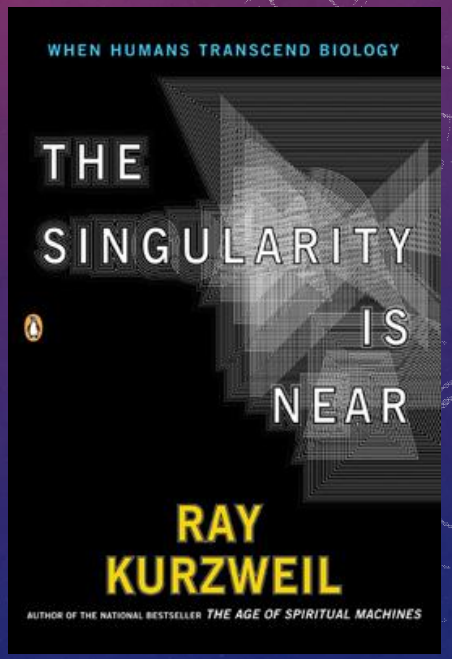
S1. Marketing Dept., staff of 45, $9,500 Travel Budget.

Department	Department Size	Travel Budget
11 25 28 31 42	1 2 3 4 5	A B C D E
○ ○ ○ ○ ○	○ ○ ○ ○ ○	○ ○ ○ ○ ○

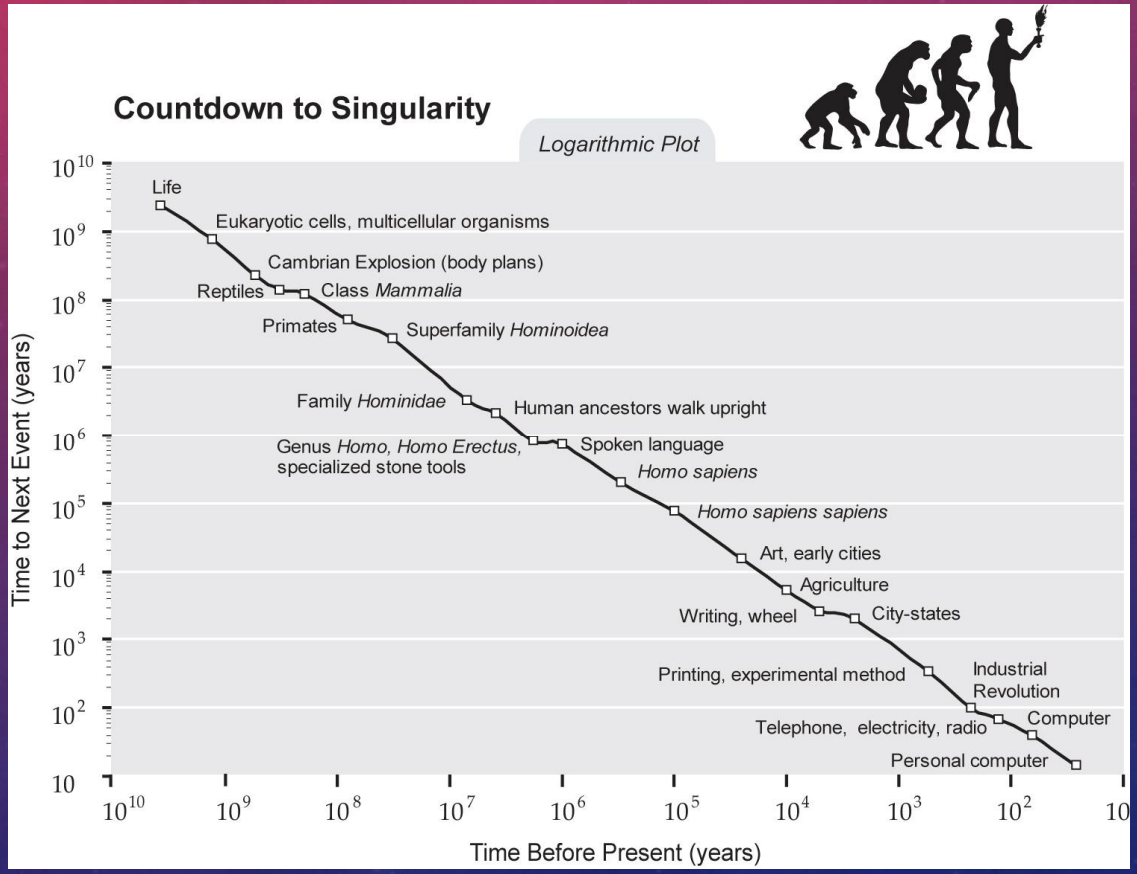
ARTIFICIAL INTELLIGENCE – SUPERINTELLIGENCE

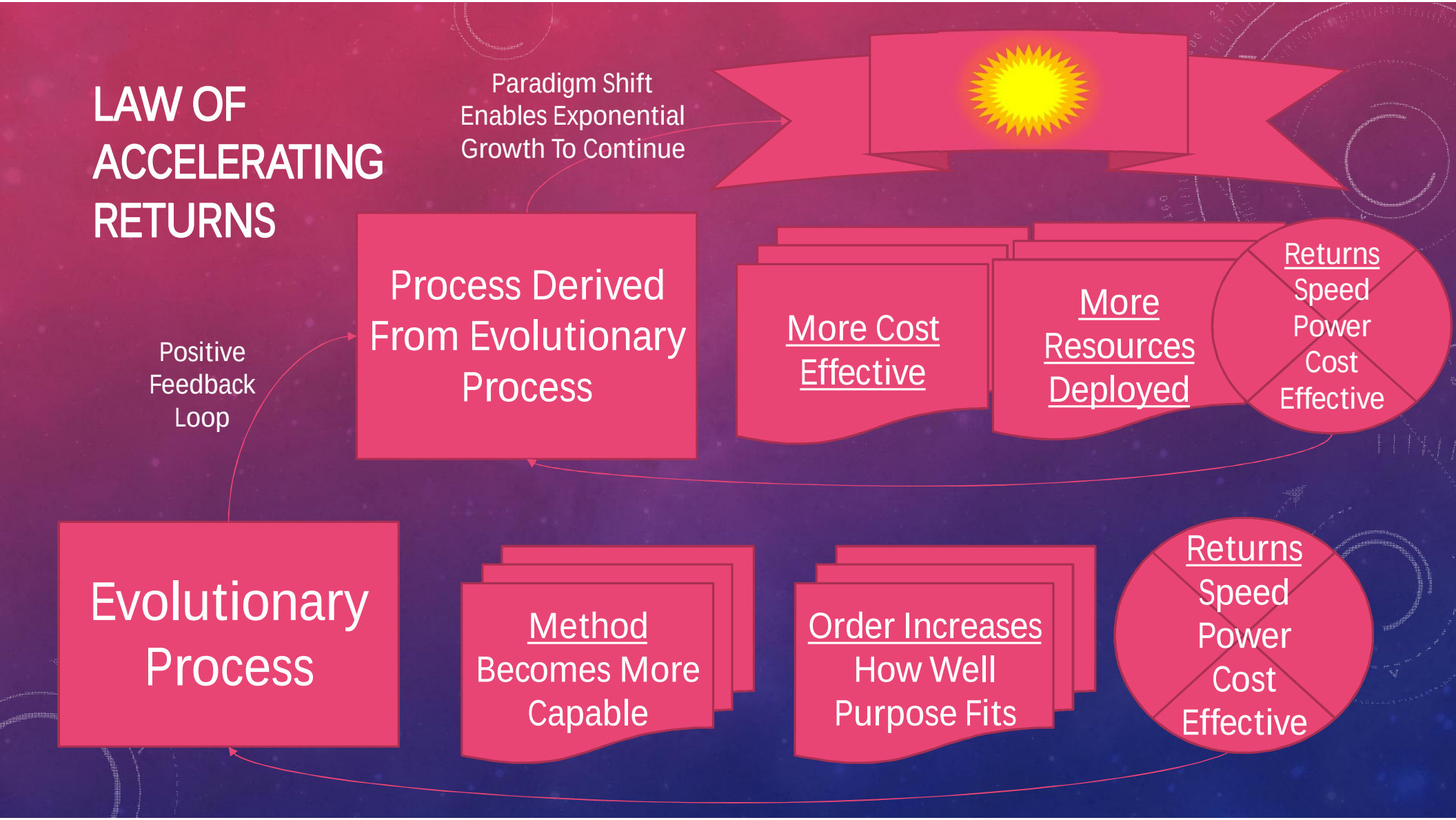
ARTIFICIAL INTELLIGENCE – RAY KURZWEIL



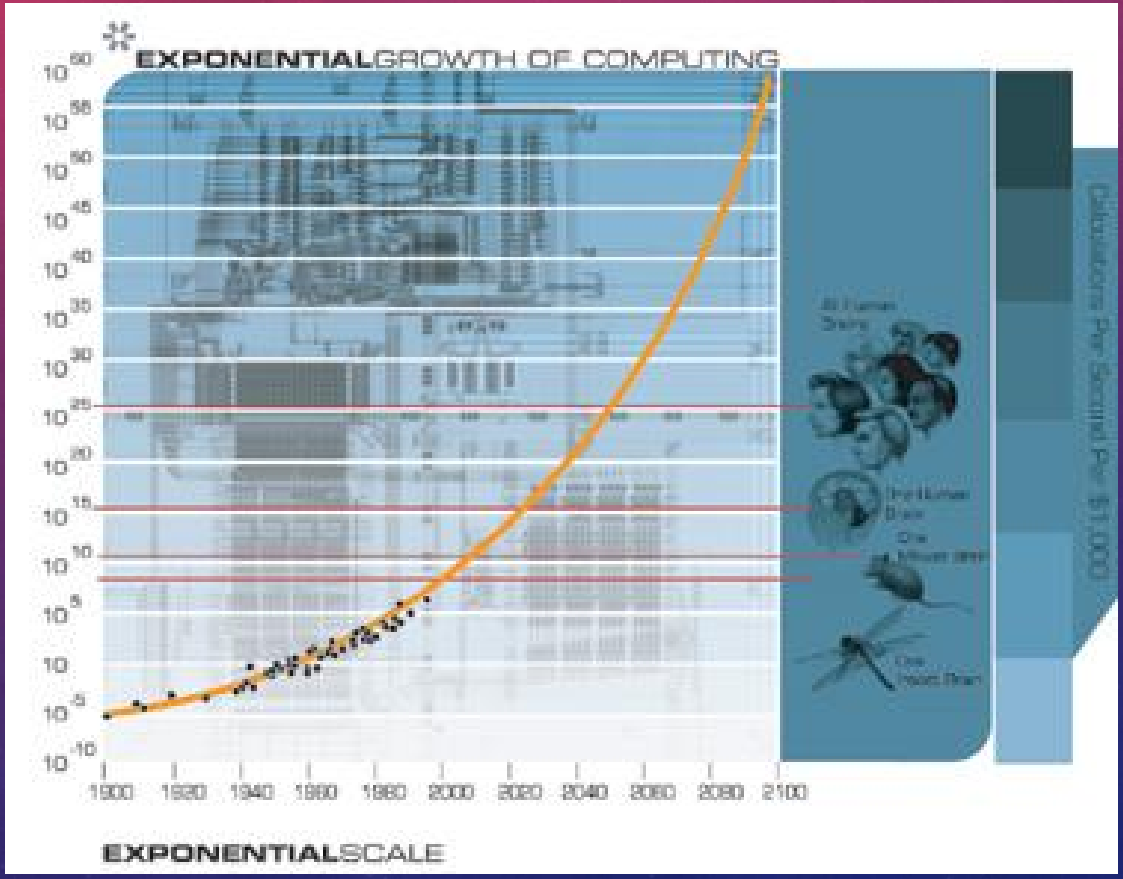


RAY KURZWEIL - SINGULARITY

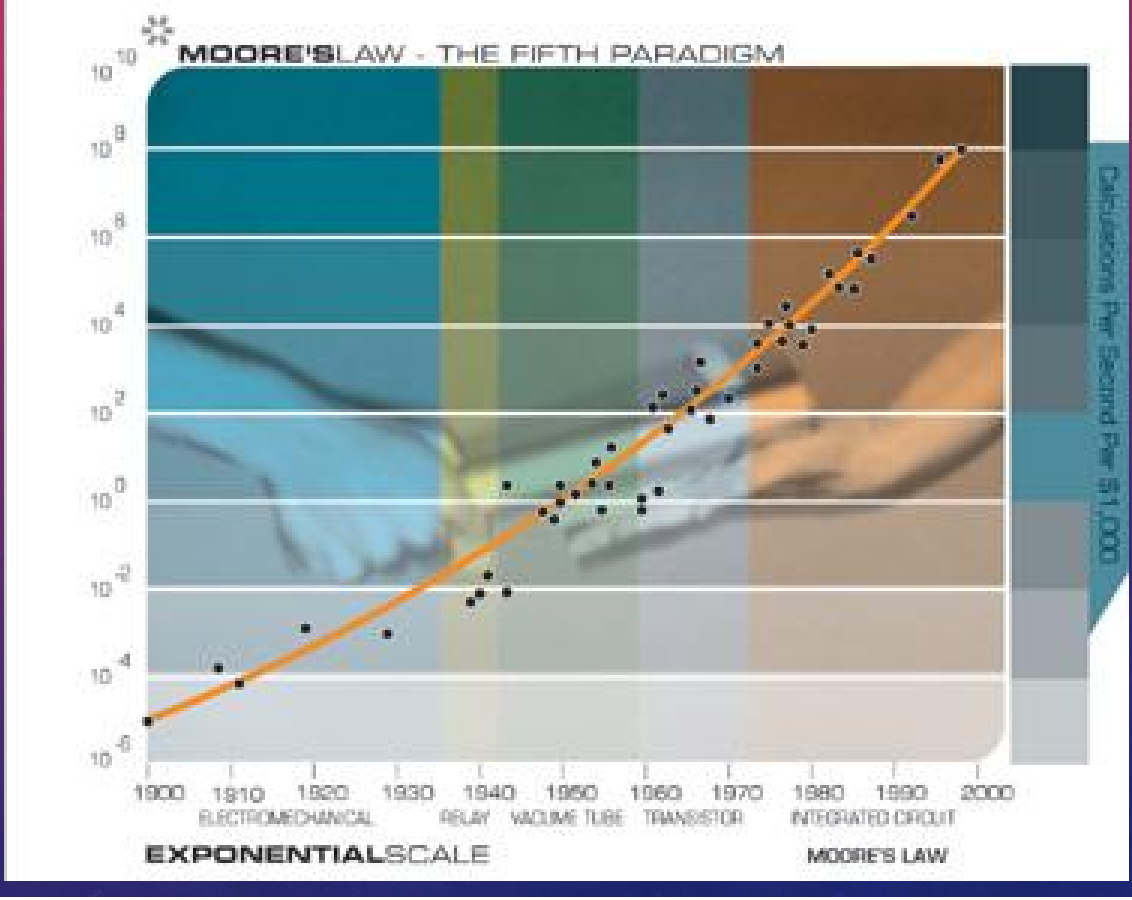



LAW OF ACCELERATING RETURNS
Paradigm Shift Enables Exponential Growth To Continue
Process Derived From Evolutionary Process
More Cost Effective
More Resources Deployed
Returns Speed Power Cost Effective
Positive Feedback Loop
Evolutionary Process
Method Becomes More Capable
Order Increases How Well Purpose Fits
Returns Speed Power Cost Effective

RAY KURZWEIL – LAW OF ACCELERATING RETURNS



RAY KURZWEIL –MOORE'S LAW



ARTIFICIAL INTELLIGENCE – ELIEZER YUDKOWSKY

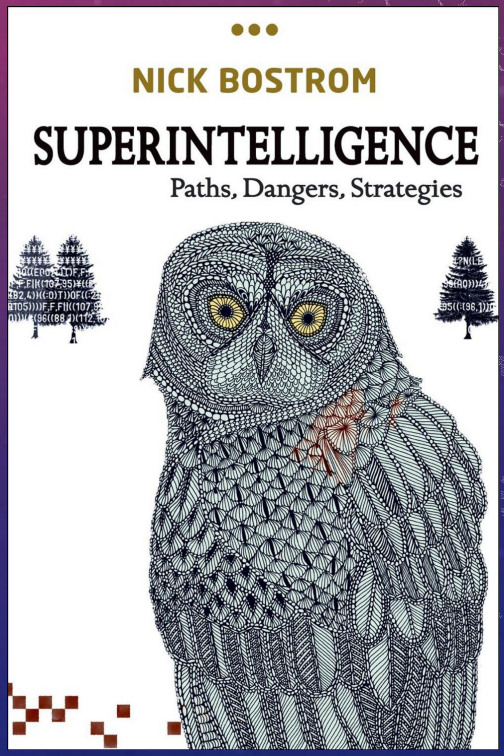
MIRI

MACHINE INTELLIGENCE RESEARCH INSTITUTE

ARTIFICIAL INTELLIGENCE – ELIEZER YUDKOWSKY

ARTIFICIAL INTELLIGENCE – COHERENT EXTRAPOLATED VOLITION (CEV)

ARTIFICIAL INTELLIGENCE – NICK BOSTROM



NICK BOSTROM – EXISTENTIAL RISK CLASSIFICATION

Extinction

Human / No Posthuman

Undesirable

Unrealized Potential

NICK BOSTROM – EXISTENTIAL RISK CLASSIFICATION

Bangs

- Misuse of Nanotechnology
- Nuclear Holocaust
- Simulation
- Genetic Engineering
- Nanotechnology Accident
- Physics Disaster
- Naturally Occurring Disease
- Asteroid / Comet
- Global Warming

Crunches

- Resource Depletion
- Misguided World Government
- Technological Arrest

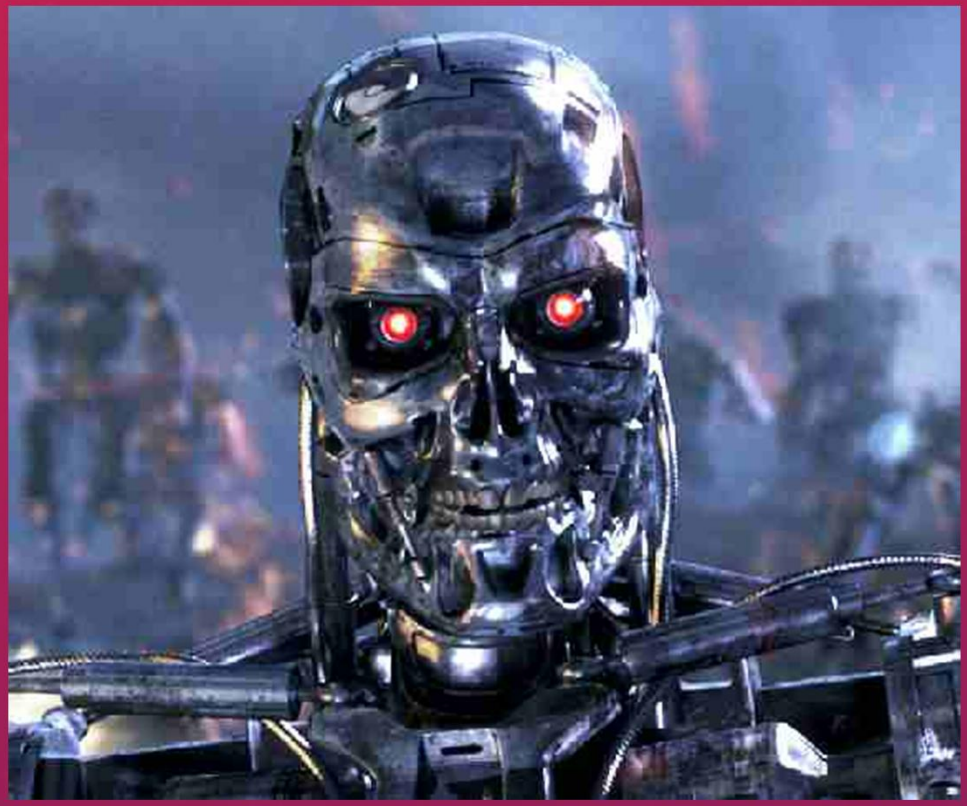
Shrieks

- Transcending Upload Takeover
- Flawed ASI
- Oppressive Regime

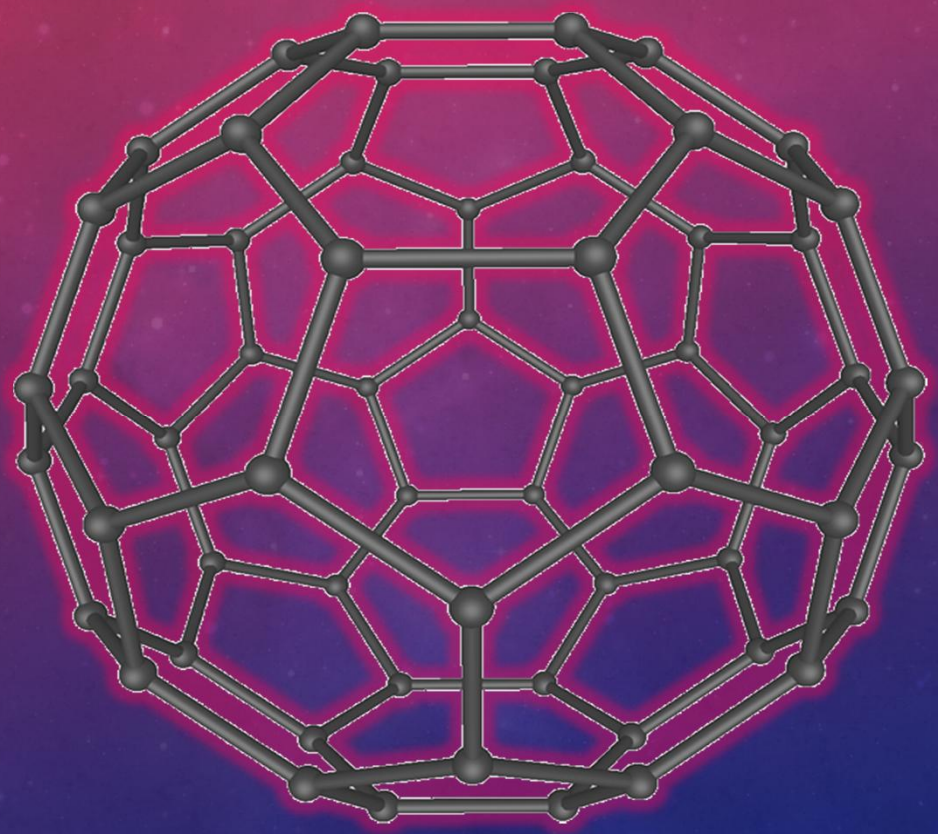
Whimpers

- Potential Eroded By Evolution
- Killed By Aliens

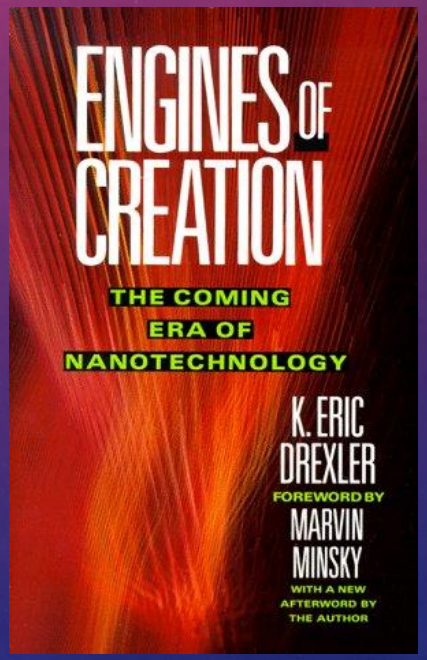
NICK BOSTROM - RECALCITRANCE

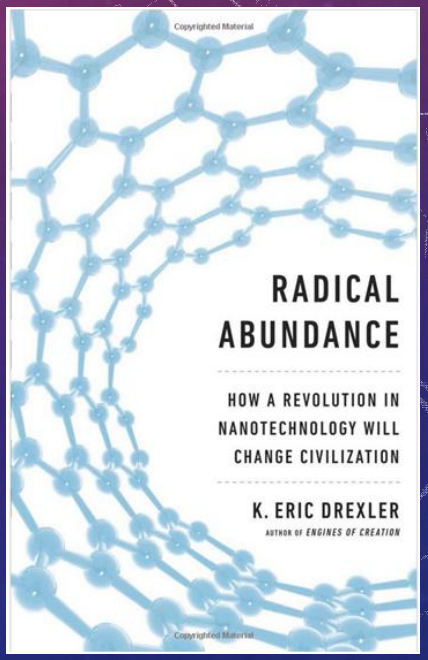


NANOTECHNOLOGY

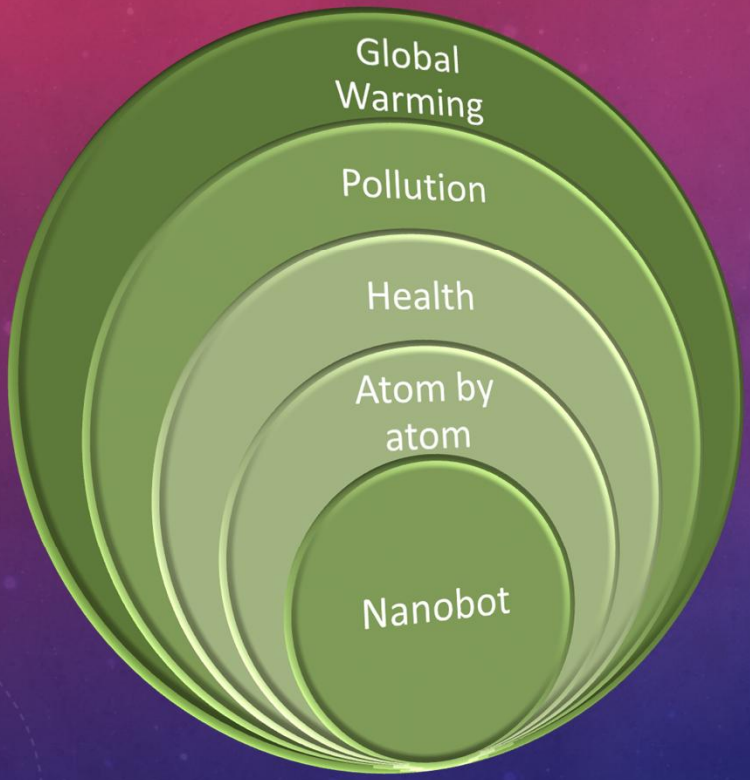


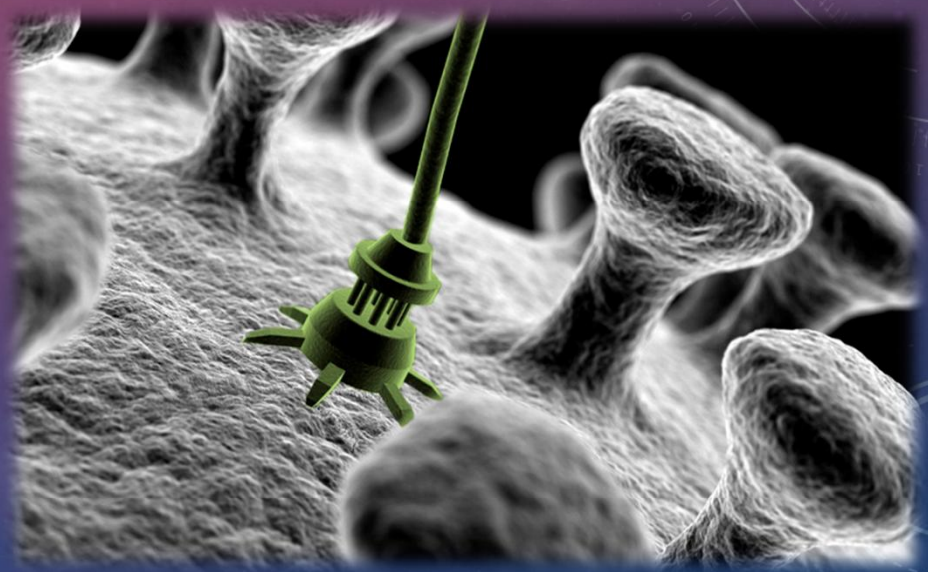
NANOTECHNOLOGY – K. ERIC DREXLER



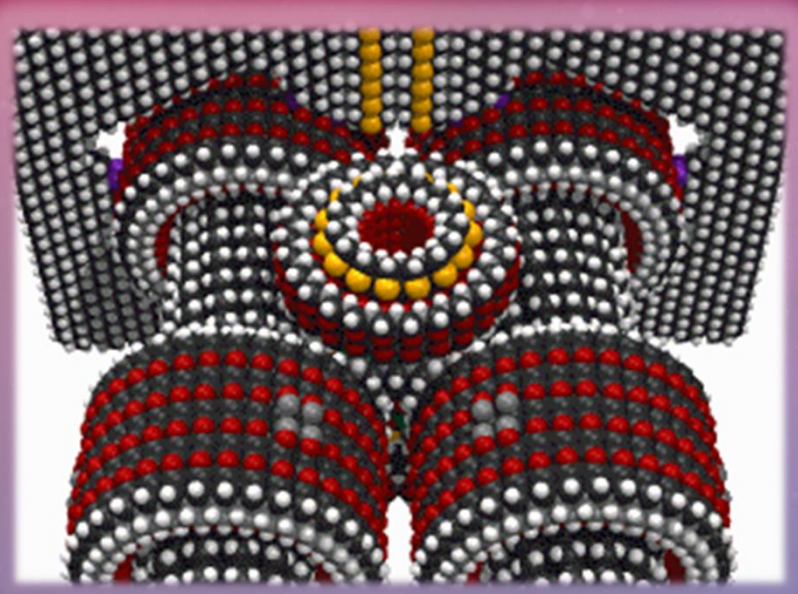


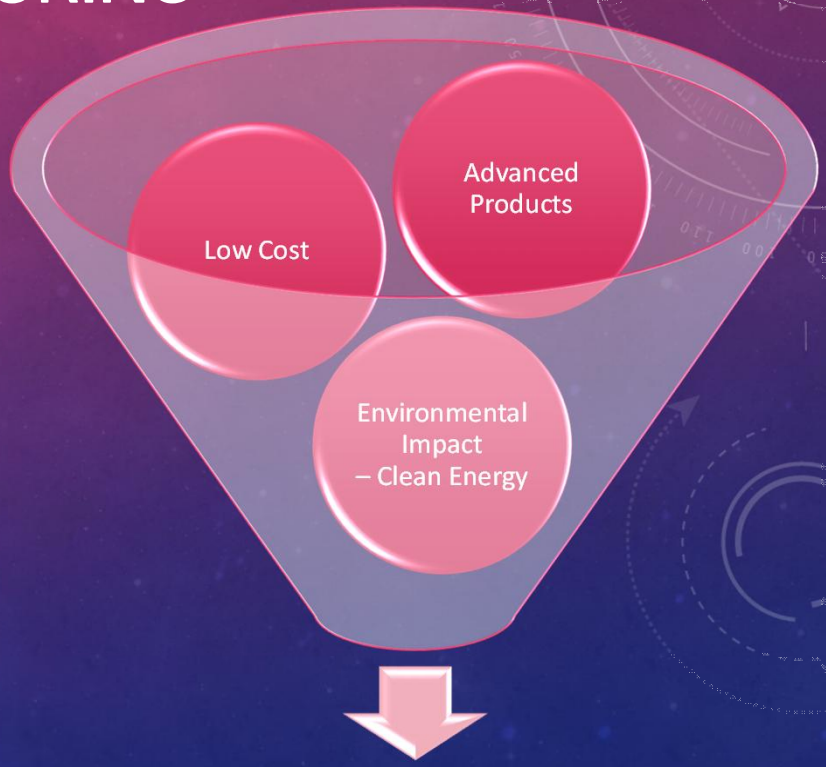
K. ERIC DREXLER
ENGINES OF CREATION





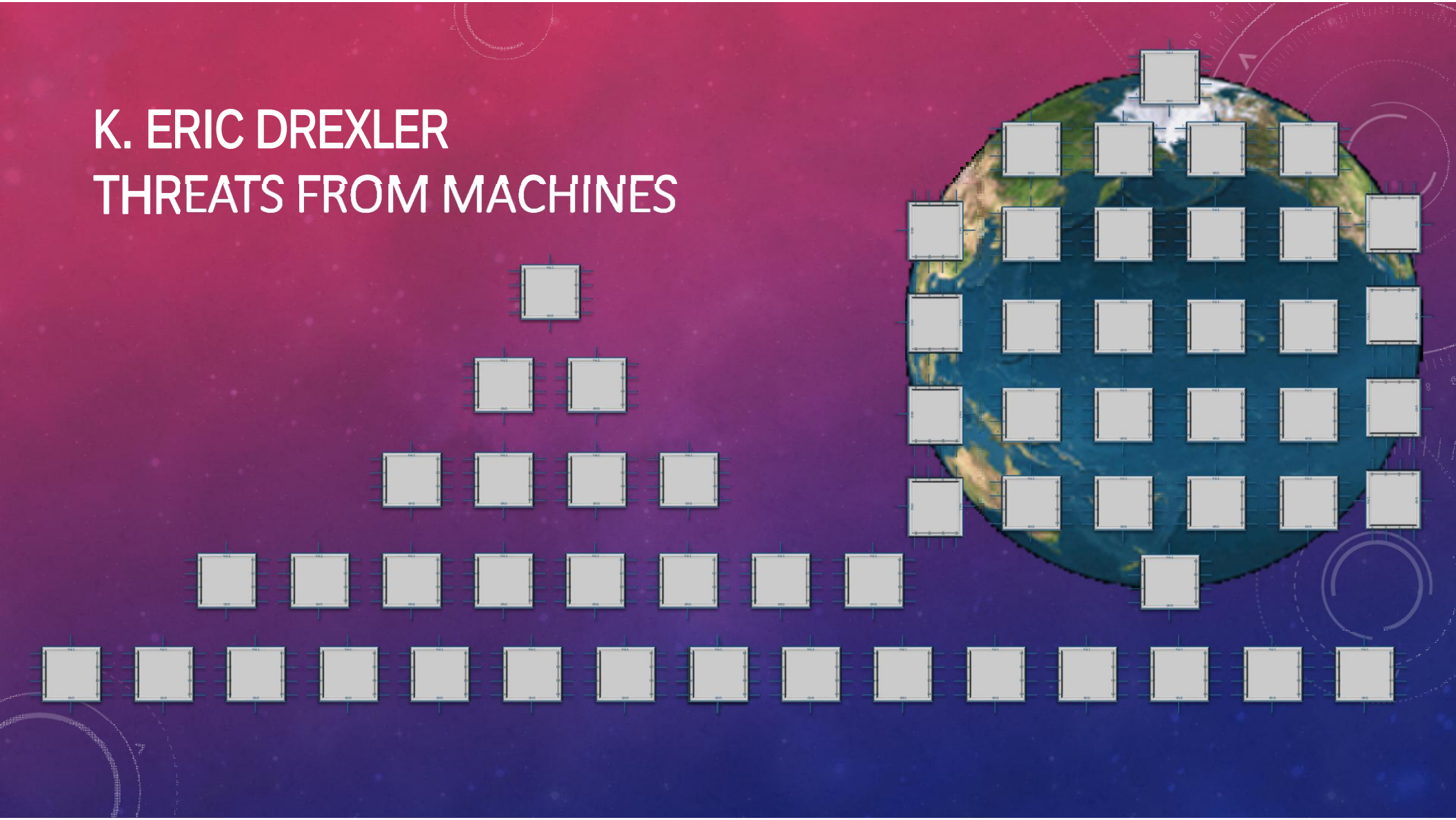
K. ERIC DREXLER
ATOMICALLY PRECISE MANUFACTURING



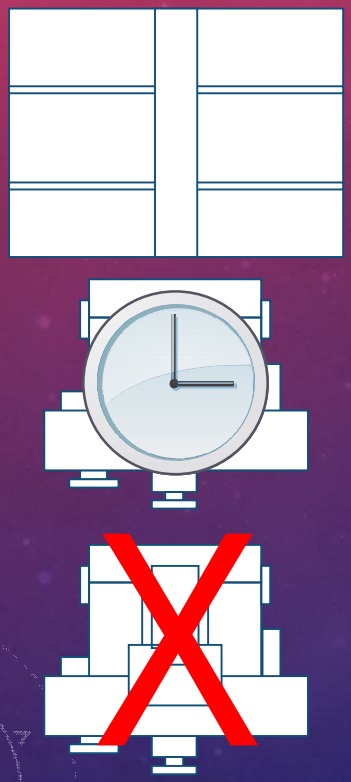


End of Industrial Revolution?

K. ERIC DREXLER
THREATS FROM MACHINES



K. ERIC DREXLER
THREATS FROM ASSEMBLER BREAKTHROUGH



K. ERIC DREXLER
PROBABILITY OF SUCCESS

Success is probable based on our experiences

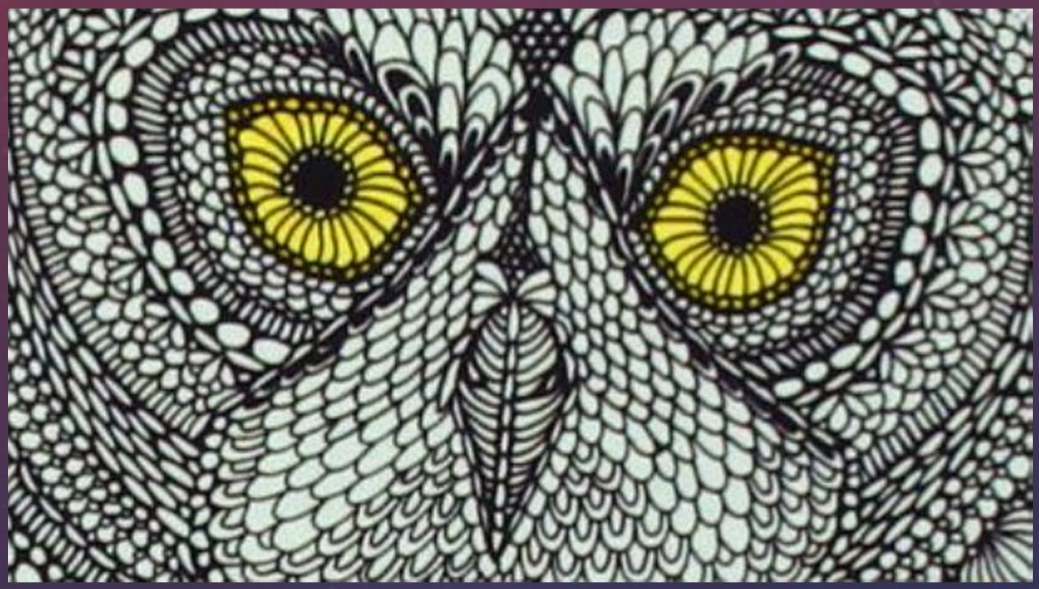
Replicators are physical systems

Biological life has been defending against invasive replicators since creation

AI can become a Nanomanager

AI – SW	Nanotechnology - HW

THREAT MODELING EXISTENTIAL RISK: NICK BOSTROM - SUPERINTELLIGENCE



https://github.com/hello7s/Existential-Threat-Models