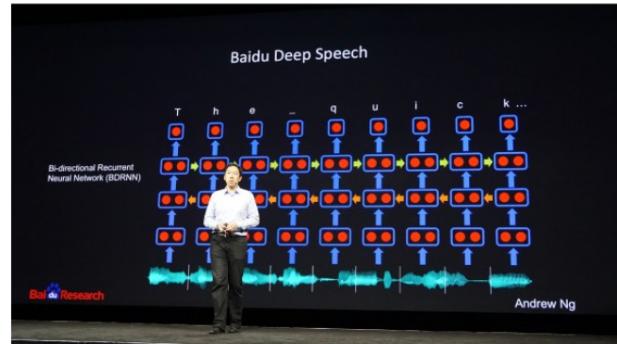
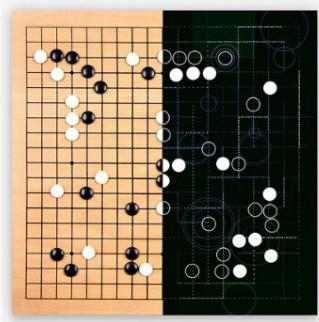


Introduction to Artificial Intelligence

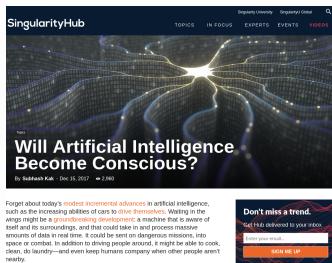
Lecture 10: Artificial General Intelligence

From technological breakthroughs...



Credits: Andrej Karpathy, Where will AGI come from?

... to popular media



SingularityHub TOPICS IN FOCUS EXPERTS EVENTS MEDIA

Will Artificial Intelligence Become Conscious?

By Subrahmanian Ramamurti | Dec. 14, 2017 | 2:30 PM

Forget about today's modest incremental advances in artificial intelligence, such as the increasing abilities of cars to drive themselves. Waiting in the wings is a revolution in AI that will change the way we live, the way we work, and the way we interact with our environment, and that could take in and process massive amounts of data in real time. It could be sent on dangerous missions to outer space, or it could be used to cook dinner, clean up after you, do laundry—and even keep humans company when other people aren't nearby.

The artificial-intelligence expert is on a mission to do it by manufacturing, starting with partners like Foxconn.

by Rachel Metz - December 14, 2017

Joining the discussion on bringing the Singularity closer to reality, Andrew Ng, co-founder of Google Brain, has joined the SingularityHub community.

Don't miss a trend.
Get Hub delivered to your inbox.
[SIGN UP](#)



MIT Technology Review Intelligent Machines

Andrew Ng Says Factories Are AI's Next Frontier



The artificial-intelligence expert is on a mission to do it by manufacturing, starting with partners like Foxconn.

by Rachel Metz - December 14, 2017

Joining the discussion on bringing the Singularity closer to reality, Andrew Ng, co-founder of Google Brain, has joined the SingularityHub community.



KILLER ROBOTS' WILL START SLAUGHTERING PEOPLE IF THEY'RE NOT BANNED SOON, AI EXPERT WARNS

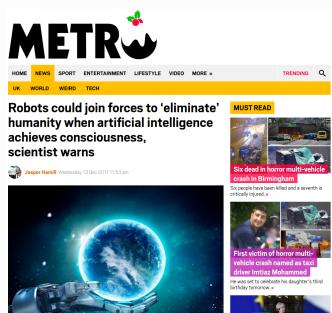
Robotics student Leslie Valente interacts with a bionic robotic hand built at the University of Regensburg in the RoboDebut exhibition at the University of Regensburg, Germany, on Dec. 14, 2017. (AP Photo/Matthias Schrader)

"These will be weapons of mass destruction"



La Tribune Science & Tech

"Si nous ne faisons rien, l'intelligence artificielle nous écrabouillera dans 30 ans"



METRO HOME NEWS SPORT ENTERTAINMENT LIFESTYLE VIDEO MORE + TRENDING

Robots could join forces to 'eliminate' humanity when artificial intelligence achieves consciousness, scientist warns

by Jasper Hand - Wednesday, 13 Dec 2017 11:53 am

MUST READ

- Six dead in horror multi-vehicle accident**
- First section of horror multi-vehicle accident in which driver imitated Mohamed**



Six people have been killed and a seventh is missing after a multi-vehicle accident in which a driver imitated Mohamed



The New York Times Home News Showbiz & TV Sport Comment Finance Travel Entertainment Life & Style UK World Politics Home Weather History Obituaries Sunday SCIENCE

Will Robots Take Our Children's Jobs?

by Alex Wellerstein - DEC. 14, 2017

Rise of the machines: Super intelligent robots could 'spell the end of the human race'

ROBOT could become conscious, turn against their masters and overthrow humanity.

by Matt Dray - DEC. 14, 2017 UPDATED DEC. 15, 2017



Stephen Hawking, the world's most famous living scientist, has warned that superintelligent robots could pose a threat to humanity.

Latest videos

- FBI Agent: 'I'm a terrorist' - US agent describes himself as 'a terrorist' in video
- US Navy officer: US Navy describes itself as 'a terrorist' in video
- Former FBI Agent: 'I'm a terrorist' - US agent describes himself as 'a terrorist' in video
- Stephen Hawking: Stephen Hawking has predicted that AI will be the terminator for human race
- Donald Trump: Donald Trump has predicted that AI will be the terminator for human race



Mail Online Science & Tech

Artificially intelligent robots could soon gain consciousness and rebel against humans to 'ELIMINATE US', scientist warns

Many scientists believe the AI robots could soon gain consciousness. They could allow them to complete dangerous missions in combat or in space. But it could also swing the other way, and lead to killer robots. In a new article for The Conversation, Professor Subrahmanian Karki explains the possible consequences if robots gain consciousness.

By SUBRAHMANYAM KARKI - DEC. 14, 2017 UPDATED DEC. 15, 2017

1.3K

From saving developing children to leading the world champion to AI robots are slowly but surely developing more and more advanced capabilities.

And many scientists, including Professor Stephen Hawking, suggest it may only be a matter of time before machines gain consciousness.

In a new article for The Conversation, Professor Subrahmanian Karki, Regents Professor of Electrical and Computer Engineering at Georgia Institute of Technology, explains the possible consequences if artificial intelligence gains consciousness.



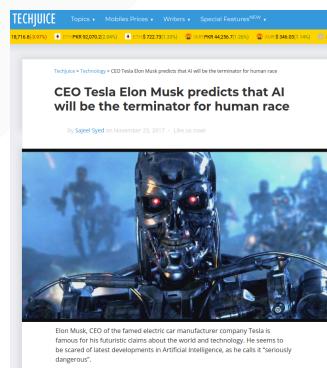
Le Figaro Actualités Start-up Tests Technologie Pratique Jeux vidéos

Pour Elon Musk, l'intelligence artificielle pourrait menacer la civilisation

by Le Figaro Staff - Mercredi 14 Décembre 2017 à 12:00 | Publié le 14 Décembre 2017 à 12:00



Elon Musk, CEO of the electric car manufacturer company Tesla is famous for his futuristic claims about the world and technology. He seems to be scared of latest developments in Artificial Intelligence, as he calls it "seriously dangerous".



TechJuice TOPICS Mobile Phones Tablets Digital Gadgets

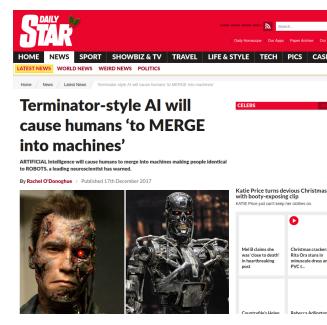
CEO Tesla Elon Musk predicts that AI will be the terminator for human race

by Sajid Syed - November 03, 2017 - 1 like on notes

Terminator-style AI will cause humans to MERGE into machines'



Elon Musk, CEO of the electric car manufacturer company Tesla is famous for his futuristic claims about the world and technology. He seems to be scared of latest developments in Artificial Intelligence, as he calls it "seriously dangerous".



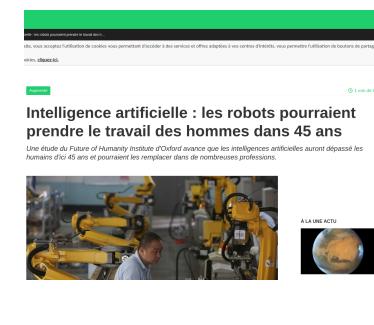
Only Star HOME NEWS SPORT SHOWBIZ & TV TRAVEL LIFE & STYLE TECH PICS CASINO

Terminator-style AI will cause humans to MERGE into machines'

by Rachel O'Donnell - Published 17/12/2017



Elon Musk, CEO of the electric car manufacturer company Tesla is famous for his futuristic claims about the world and technology. He seems to be scared of latest developments in Artificial Intelligence, as he calls it "seriously dangerous".



Dailymotion Home News Latest News

Intelligence artificielle : les robots pourraient prendre le travail des hommes dans 45 ans

Une survie de l'humanité humaine pourrait dépendre que les intelligences artificielles auront dépassé les humains d'ici 45 ans et pourraient les remplacer dans de nombreuses professions.



A LA UNE ACTU

Artificial narrow intelligence

- Artificial intelligence today is still very **narrow**.
 - Modern AI systems often reach super-human level performance.
 - ... but only at **very specific problems!**
 - They **do not generalize** to the real world nor to arbitrary tasks.

AlphaGo

Convenient properties of AlphaGo:

- Deterministic (no noise in the game).
- Fully observed (each player has complete information)
- Discrete action space (finite number of actions possible)
- Perfect simulator (the effect of any action is known exactly)
- Short episodes (200 actions per game)
- Clear and fast evaluation (as stated by Go rules)
- Huge dataset available (games)



Credits: Andrej Karpathy, Where will AGI come from?

Picking challenge



Can we run AlphaGo on a robot for the Amazon Picking Challenge?

Picking challenge



- Deterministic: OK
- Fully observed: OKish
- Discrete action space: OK
- Perfect simulator: TROUBLE
- Short episodes: challenge
- Clear and fast evaluation: not good
- Huge dataset available: challenge

Artificial general intelligence

- Artificial general intelligence (AGI) is the intelligence of a machine that could successfully perform any intellectual task that a human being can.
- No clear definition, but there is an agreement that AGI is required to do the following:
 - reason, use strategy, solve puzzle,
 - make judgments under uncertainty,
 - represent knowledge, including commonsense knowledge,
 - plan,
 - learn,
 - communicate in natural language,
 - integrate all these skills towards common goals.
- This is similar to our definition of thinking rationally, but applied broadly to any set of tasks.

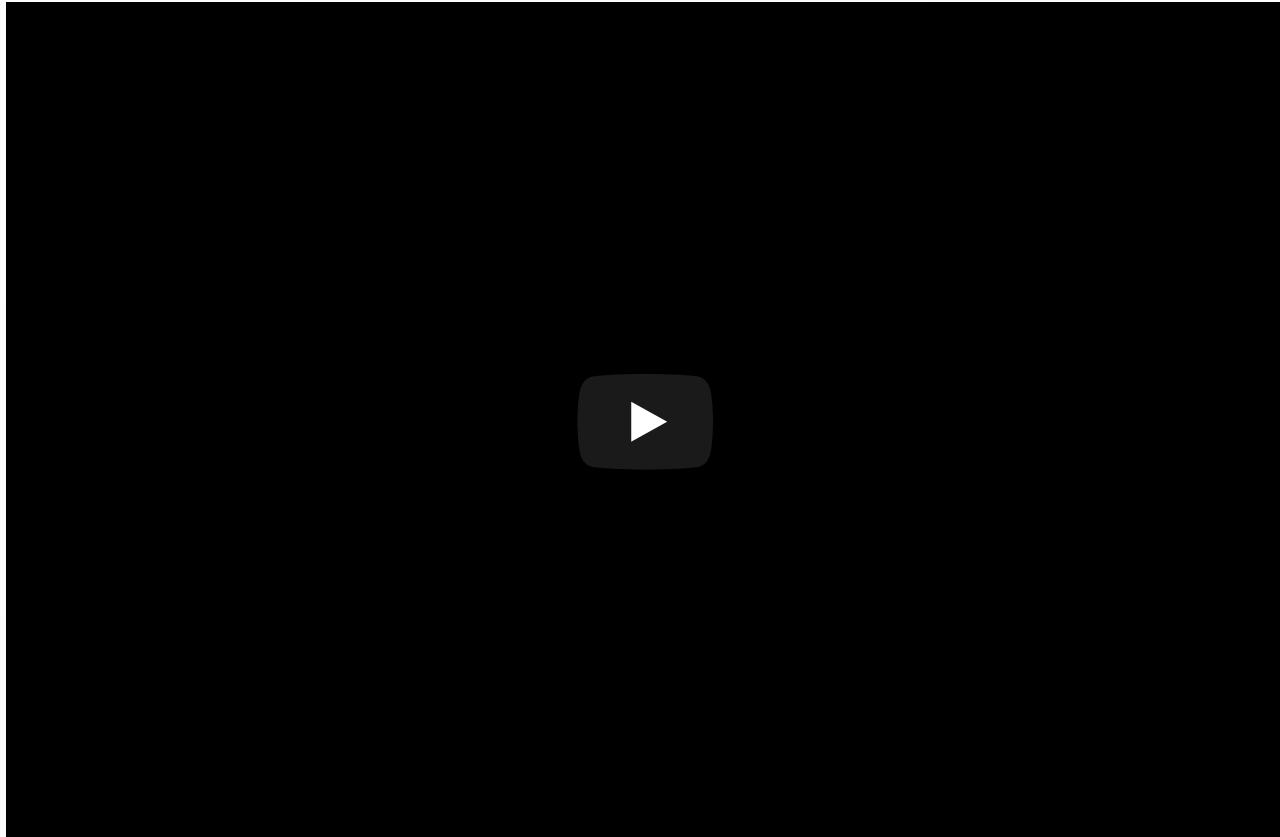
Singularity



Irving John Good (1965):

- Let an **ultraintelligent** machine be defined as a machine that can far surpass all the intellectual activities of any man however clever.
- Since the design of machines is one of these intellectual activities, an ultraintelligent machine could **design even better machines**.
- There would then unquestionably be an '**intelligence explosion**', and the intelligence of man would be left far behind.
- Thus the first ultraintelligent machine is the **last invention** that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.

Superintelligence



What happens when our computers get smarter than we are? Nick Bostrom

How to build AGI?

Several working **hypothesis**:

- **Supervised learning**: "It works, just scale up!"
- **Unsupervised learning**: "It will work, if we only scale up!"
- **AIXI**: "Guys, I can write down an equation for optimal AI."
- **Brain simulation**: "This will work one day, right?"
- **Artificial life**: "Let just do what Nature did."

Or maybe something else?

AIXI

Start with an equation

$$\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi$$

- $\Upsilon(\pi)$ formally defines the universal intelligence of an agent π .
- μ is the environment of the agent and E is the set of all computable reward bounded environments.
- $V_\mu^\pi = \mathbb{E}[\sum_{i=1}^{\infty} R_i]$ is the expected sum of future rewards when the agent π interacts with environment μ .
- $K(\cdot)$ is the Kolmogorov complexity, such that $2^{-K(\mu)}$ weights the agent's performance in each environment, inversely proportional to its complexity.
 - Intuitively, $K(\mu)$ measures the complexity of the shortest Universal Turing Machine program that describes the environment μ .

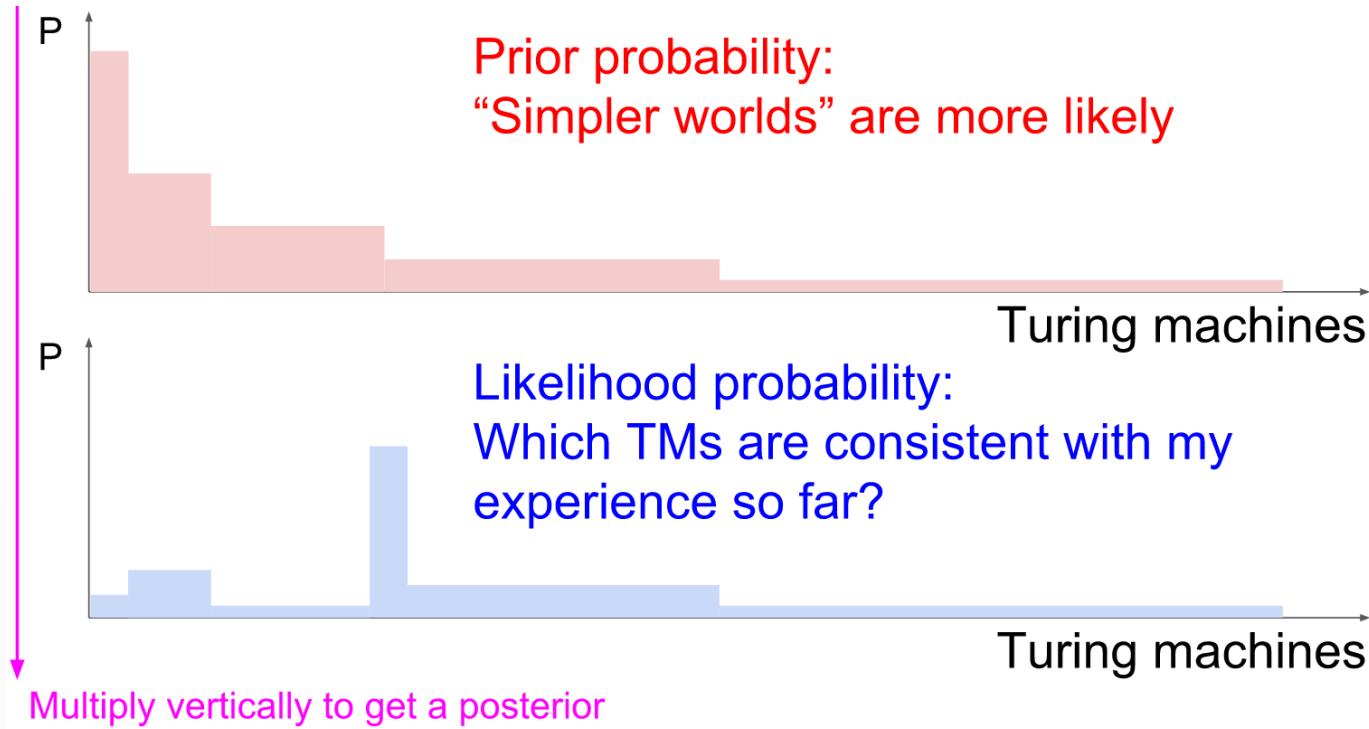
AIXI

$$\bar{\Upsilon} = \max_{\pi} \Upsilon(\pi) = \Upsilon(\pi^{AIXI})$$

π^{AIXI} is a **perfect** theoretical agent.

System identification

- Which Turing machine is the agent in? If it knew, it could plan perfectly.
- Use the [Bayes rule](#) to update the agent beliefs given its experience so far.



Acting optimally (1)

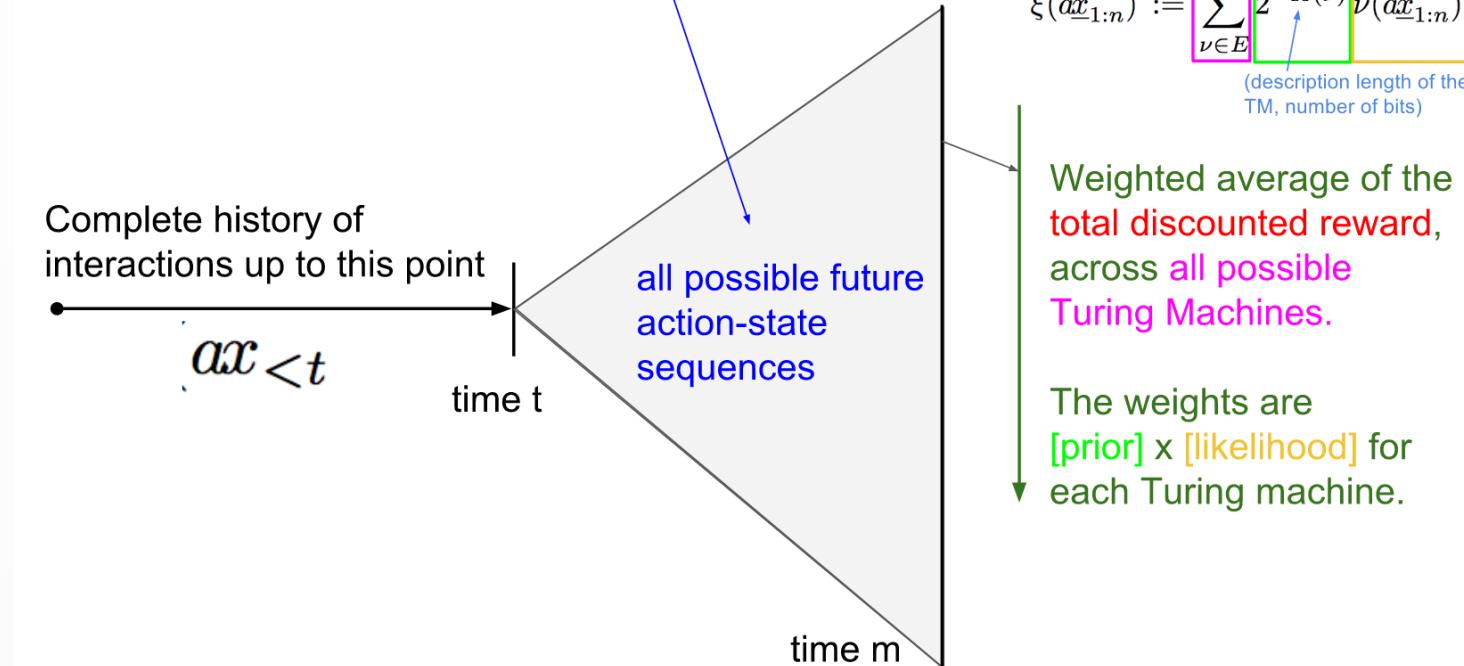
- The agent always picks the action which has the greatest expected reward.
- For every environment $\mu \in E$, the agent must:
 - Take into account how likely it is that it is facing μ given the interaction history so far, and the prior probability of μ .
 - Consider all possible future interactions that might occur, assuming optimal future actions.
 - Evaluate how likely they are.
 - Then select the action that maximizes the expected future reward.

Acting optimally (2)

$$a_t^{\pi^\xi} := \arg \max_{a_t} \lim_{m \rightarrow \infty} \sum_{x_t} \max_{a_{t+1}} \sum_{x_{t+1}} \cdots \max_{a_m} \sum_{x_m} [\gamma_t r_t + \cdots + \gamma_m r_m] \xi(ax_{<t} ax_{t:m})$$

$$\xi(ax_{1:n}) := \sum_{\nu \in E} 2^{-K(\nu)} \nu(ax_{1:n})$$

(description length of the TM, number of bits)

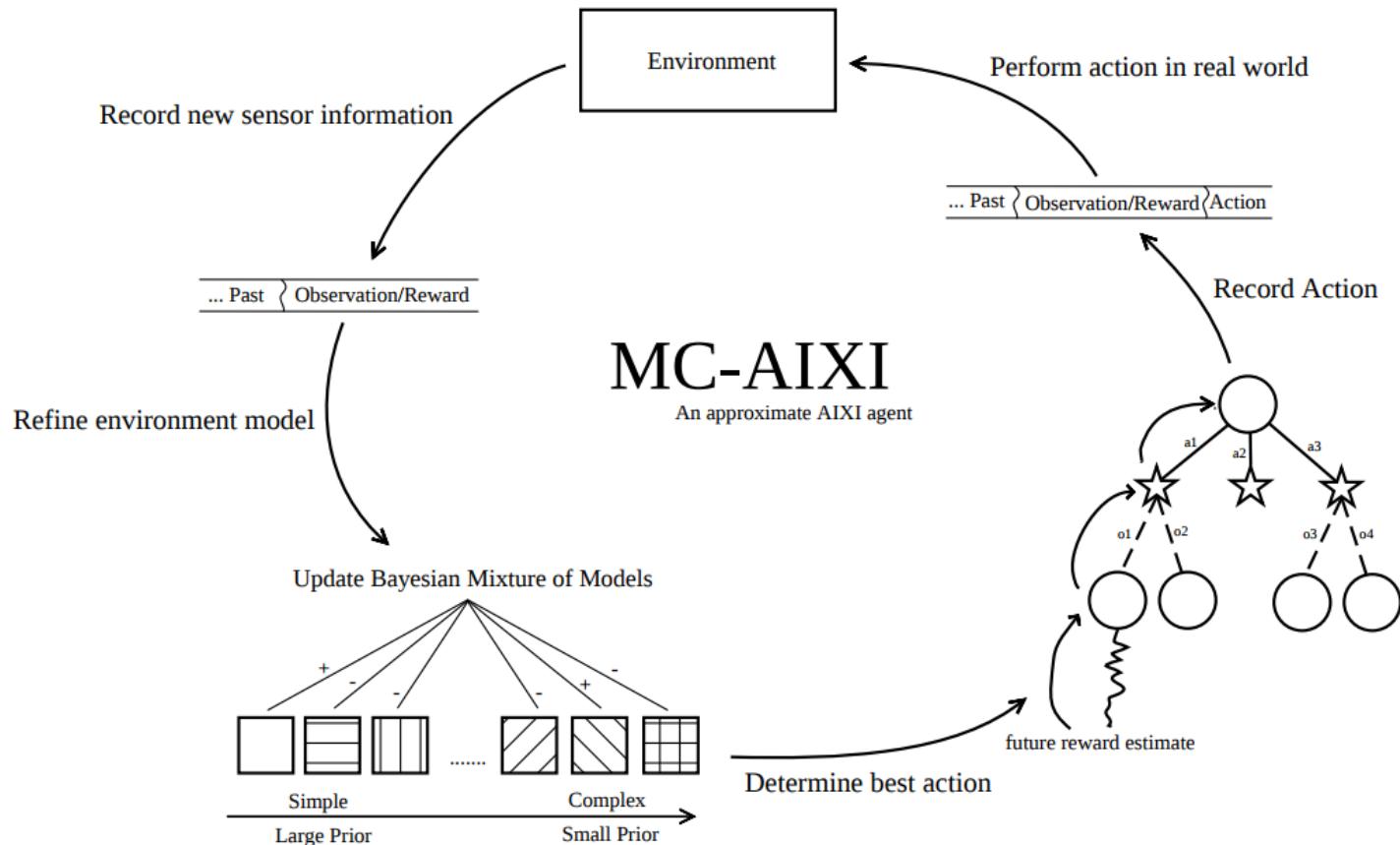


Incomputability

$$a_t^{\pi^\xi} := \arg \max_{a_t} \lim_{m \rightarrow \infty} \left[\sum_{x_t} \max_{a_{t+1}} \sum_{x_{t+1}} \cdots \max_{a_m} \sum_{x_m} [\gamma_t r_t + \dots + \gamma_m r_m] \xi(\underline{ax}_{<t} \underline{ax}_{t:m}) \right]$$

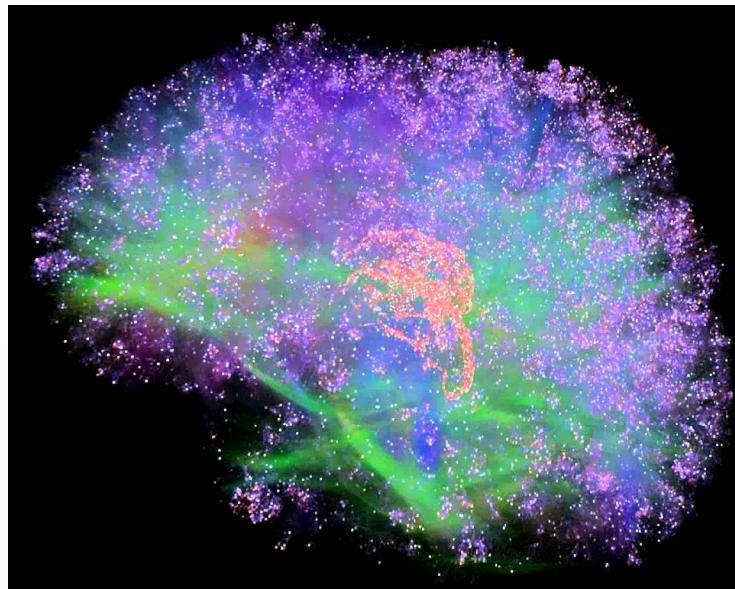
$$\xi(\underline{ax}_{1:n}) := \sum_{\nu \in E} 2^{-K(\nu)} \nu(\underline{ax}_{1:n})$$

Monte Carlo approximation



Brain simulation

Whole brain emulation



- A hypothesis for AGI is **whole brain simulation**.
 - A low-level brain model is built by scanning and mapping a biological brain in detail and copying its state into a computer system.
 - The simulation is **so faithful** that it would behave in the same way as the original.
 - Therefore, the computer-run model would be as intelligent.
- Initiatives: Blue Brain Project, Human Brain Project, Neuralink, etc.

Obstacles

- How to **measure** a complete brain state?
- At what level of abstraction?
- How to model the dynamics?
- How do you simulate the environment to feed into senses?
- Various **ethical dilemmas**.

Mind upload

- Hypothetically, whole brain emulation would enable mind upload.
 - The mental state of a particular brain substrate could be scanned and copied into a computer.
 - The computer could then run a simulation of the brain's information processing, such that it responds in the same way as the original brain.
- That is, simulation would be indistinguishable from reality.

ARE YOU LIVING IN A COMPUTER SIMULATION?

BY NICK BOSTROM

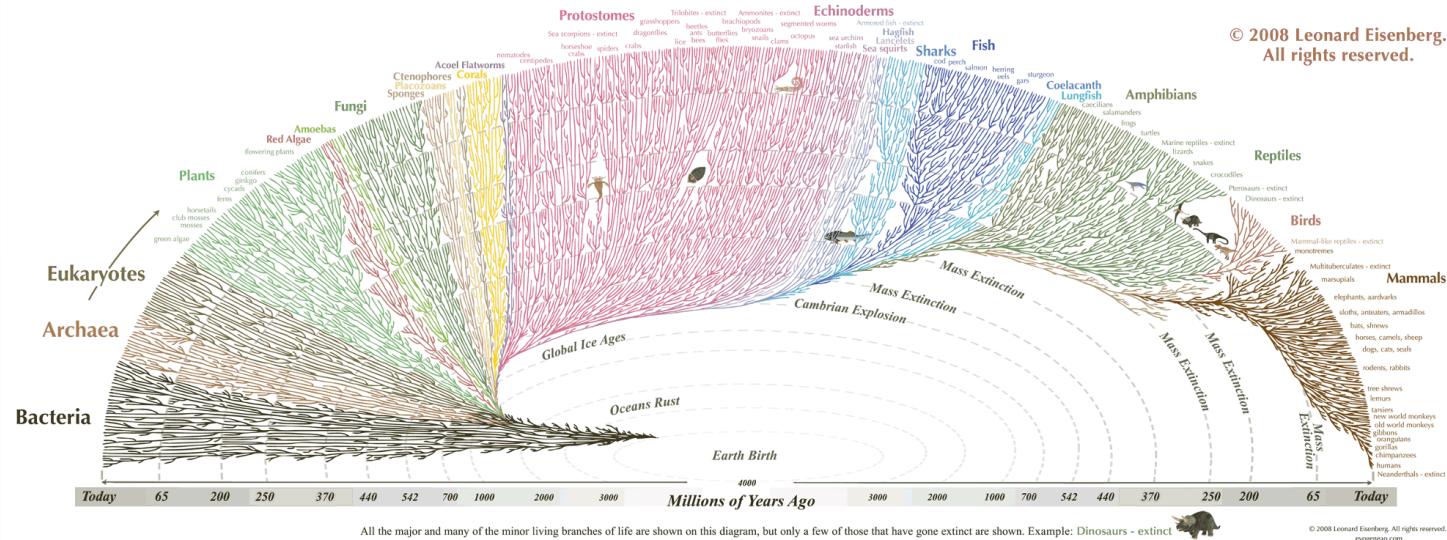
[Published in *Philosophical Quarterly* (2003) Vol. 53, No. 211, pp. 243-255. (First version: 2001)]

This paper argues that *at least one* of the following propositions is true: (1) the human species is very likely to go extinct before reaching a “posthuman” stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation. A number of other consequences of this result are also discussed.

Artificial life

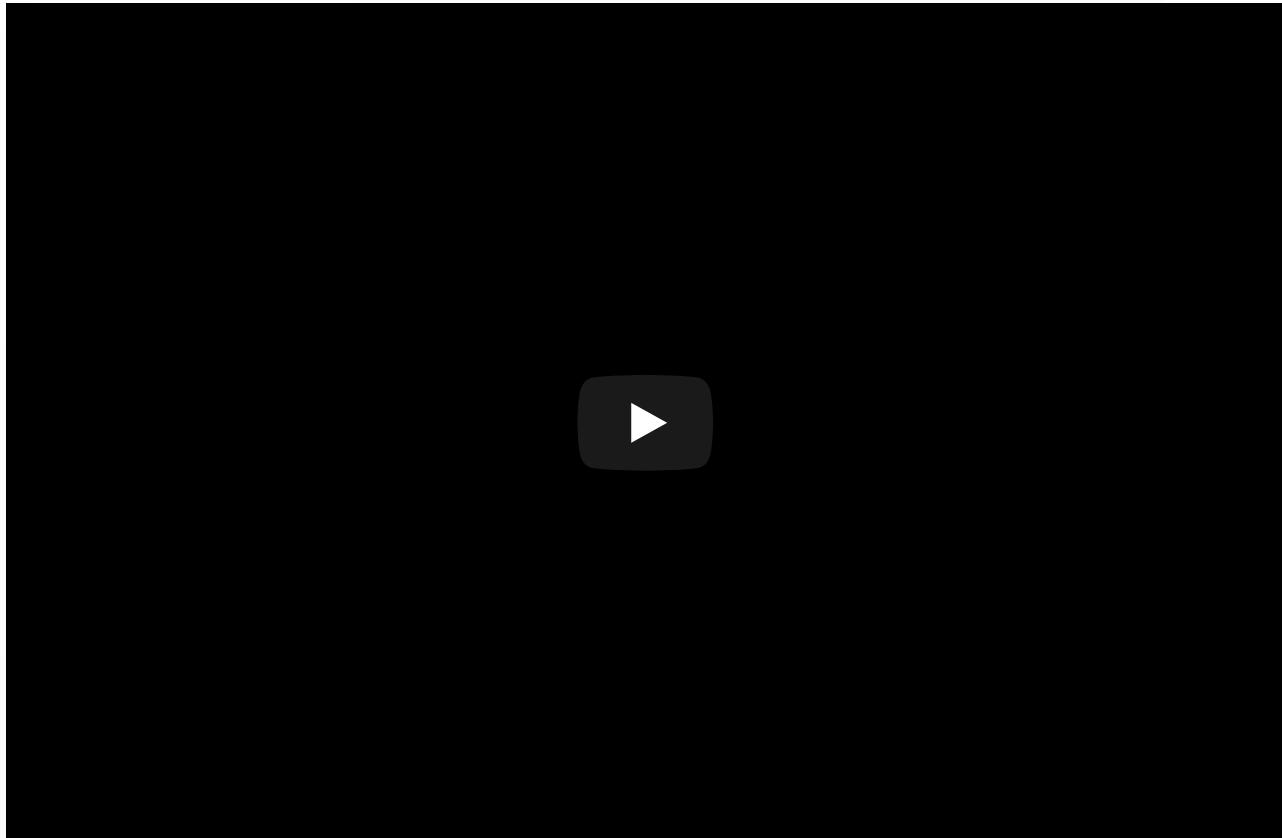
How did intelligence arise in Nature?

© 2008 Leonard Eisenberg.
All rights reserved.



Artificial life

- Artificial life is the study of systems related to natural life, its processes and its evolution, through the use of simulations with computer models, robotics or biochemistry.
- One of its goals is to synthesize life in order to understand its origins, development and organization.
- There are three main kinds of artificial life, named after their approaches:
 - Software approaches (soft)
 - Hardware approaches (hard)
 - Biochemistry approaches (wet)
- Artificial life is related to AI since synthesizing complex life forms would, hypothetically, induce intelligence.
- The field of AI has traditionally used a top down approach. Artificial life generally works from the bottom up.



The line between life and not-life. Martin Hanczyc

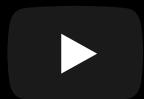
Evolution for AGI

- Evolution may **hypothetically** be interpreted as an (unknown) algorithm.
- This algorithm gave rise to AGI.
 - e.g., it induced humans.
- Can we **simulate** the **evolutionary process** to reproduce life and intelligence?
- Note that using software simulation, we can work at a high level of abstraction.
 - We don't have to simulate physics or chemistry to simulate evolution.
 - We can also bootstrap the system with agents that are better than random.

Evolutionary algorithms

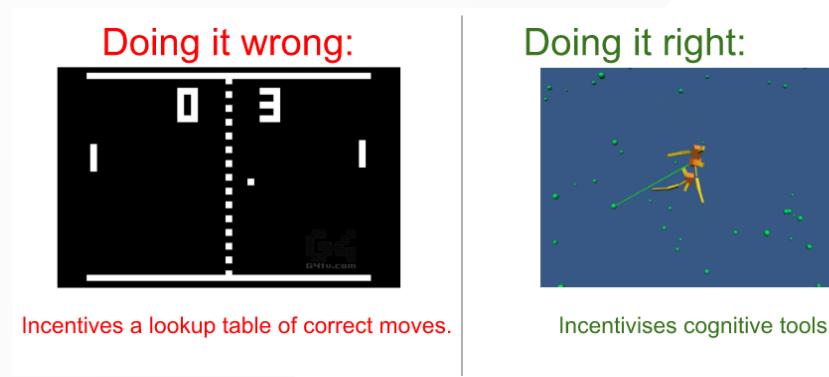
- Start with a **random population** of **creatures**.
- Each creature is **tested for their ability** to perform a given task.
 - e.g., swim in a simulated environment.
 - e.g., stay alive as long as possible (without starving or being killed).
- The **most successful survive**.
- Their virtual genes containing coded instructions for their growth are copied, combined and mutated to **make offspring** for a new population.
- The new creatures are tested again, some of which may be improvements on their parents.
- As this cycle of variation and selection continues, creatures with more and more successful behaviors may **emerge**.





Environments for AGI?

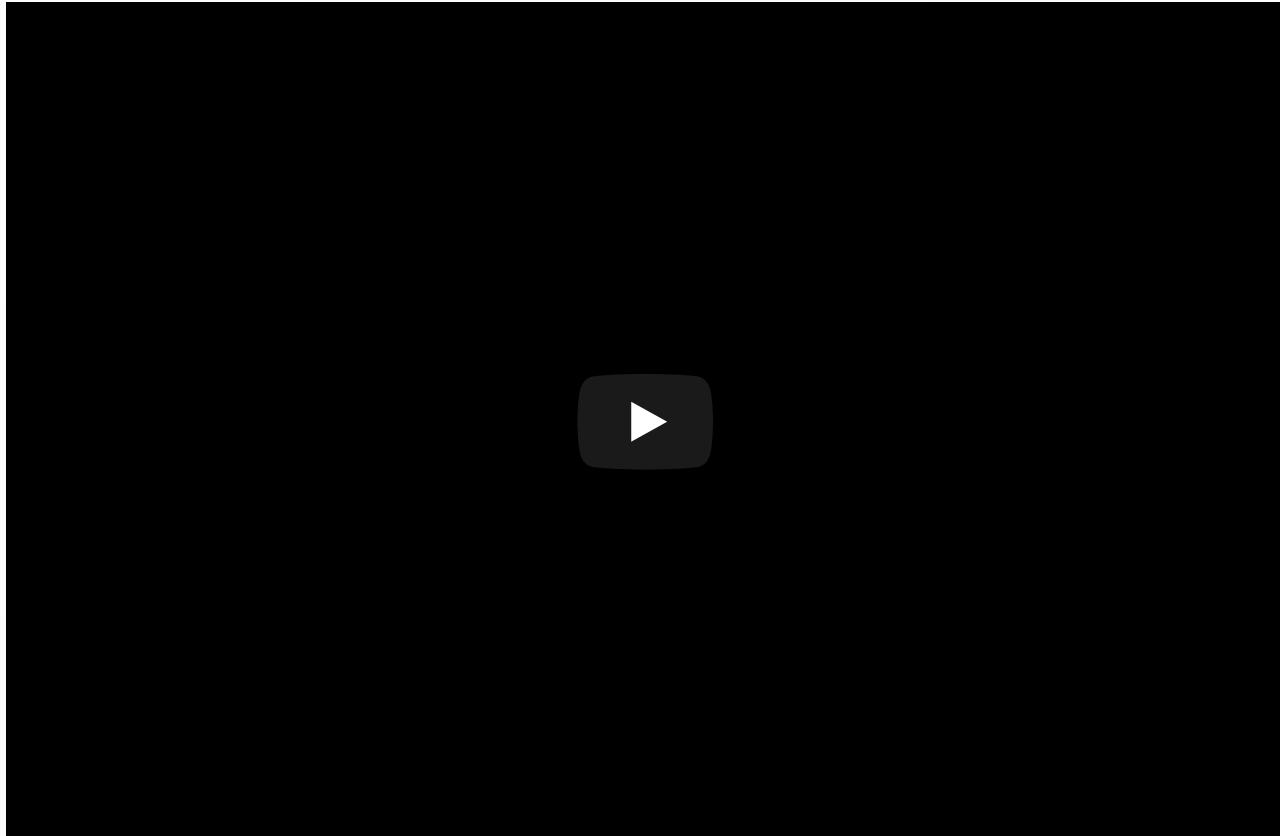
- For the emergence of generally intelligent creatures, we presumably need environments that **incentivize** the emergence of a **cognitive toolkit**.
 - attention, memory, knowledge representation, reasoning, emotions, forward simulation, skill acquisition, ...



- Multi-agent** environments are certainly better because of:
 - Variety**: the environment is parameterized by its agent population. The optimal strategy must be derived dynamically.
 - Natural curriculum**: the difficulty of the environment is determined by the skill of the other agents.

Conclusions

A note of optimism



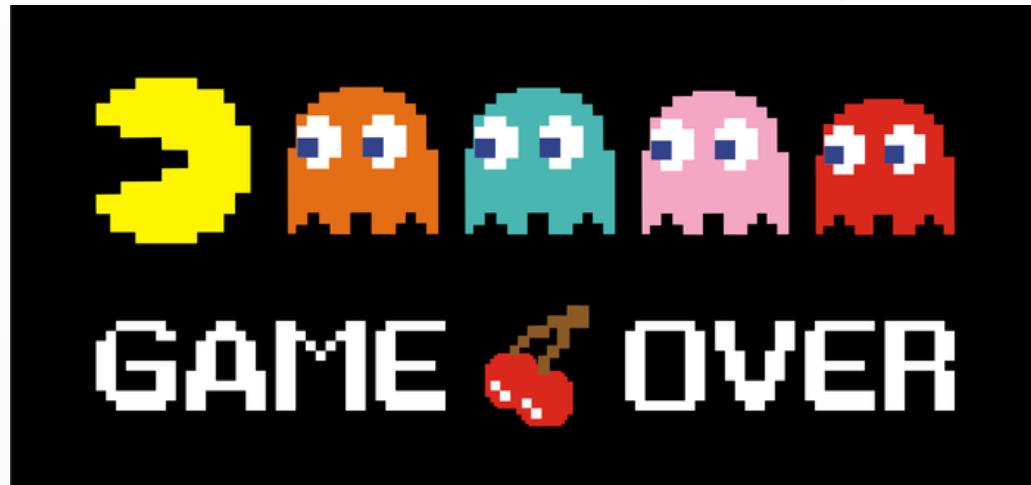
Don't fear intelligent machines, work with them. Garry Kasparov

Summary

- Lecture 1: Foundations
- Lecture 2: Solving problems by searching
- Lecture 3: Adversarial search
- Lecture 4: Constraint satisfaction problems
- Lecture 5: Representing uncertain knowledge
- Lecture 6: Inference in Bayesian networks
- Lecture 7: Reasoning over time
- Lecture 8: Learning
- Lecture 9: Communication
- Lecture 10: Artificial General Intelligence

Going further

- ELEN0062: Introduction to Machine Learning
- INFO8004: Advanced Machine Learning
- INFOXXXX: Deep Learning (Spring 2019)
- INFO8003: Optimal decision making for complex problems
- INFO0948: Introduction to Intelligent robotics
- INFO0049: Knowledge representation
- ELEN0016: Computer vision



Thanks for following Introduction to AI!

Readings

- Bostrom, Nick. *Superintelligence*. Dunod, 2017.
- Legg, Shane, and Marcus Hutter. "Universal intelligence: A definition of machine intelligence." *Minds and Machines* 17.4 (2007): 391-444.
- Hutter, Marcus. "One decade of universal artificial intelligence." *Theoretical foundations of artificial general intelligence* (2012): 67-88.
- Sims, Karl. "Evolving 3D morphology and behavior by competition." *Artificial life* 1.4 (1994): 353-372.
- Kasparov, Garry. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, 2017.