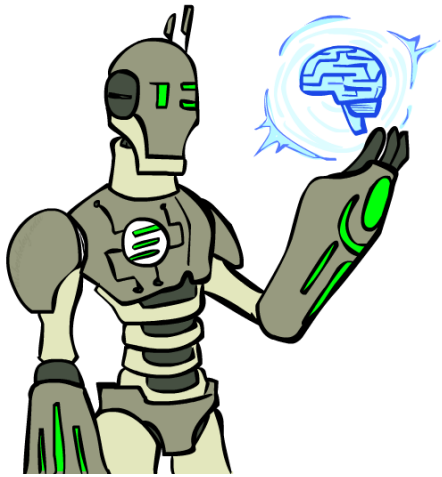
Introduction to Artificial Intelligence

Lecture 11: Artificial General Intelligence

Prof. Gilles Louppe
g.louppe@uliege.be

LIÈGE université

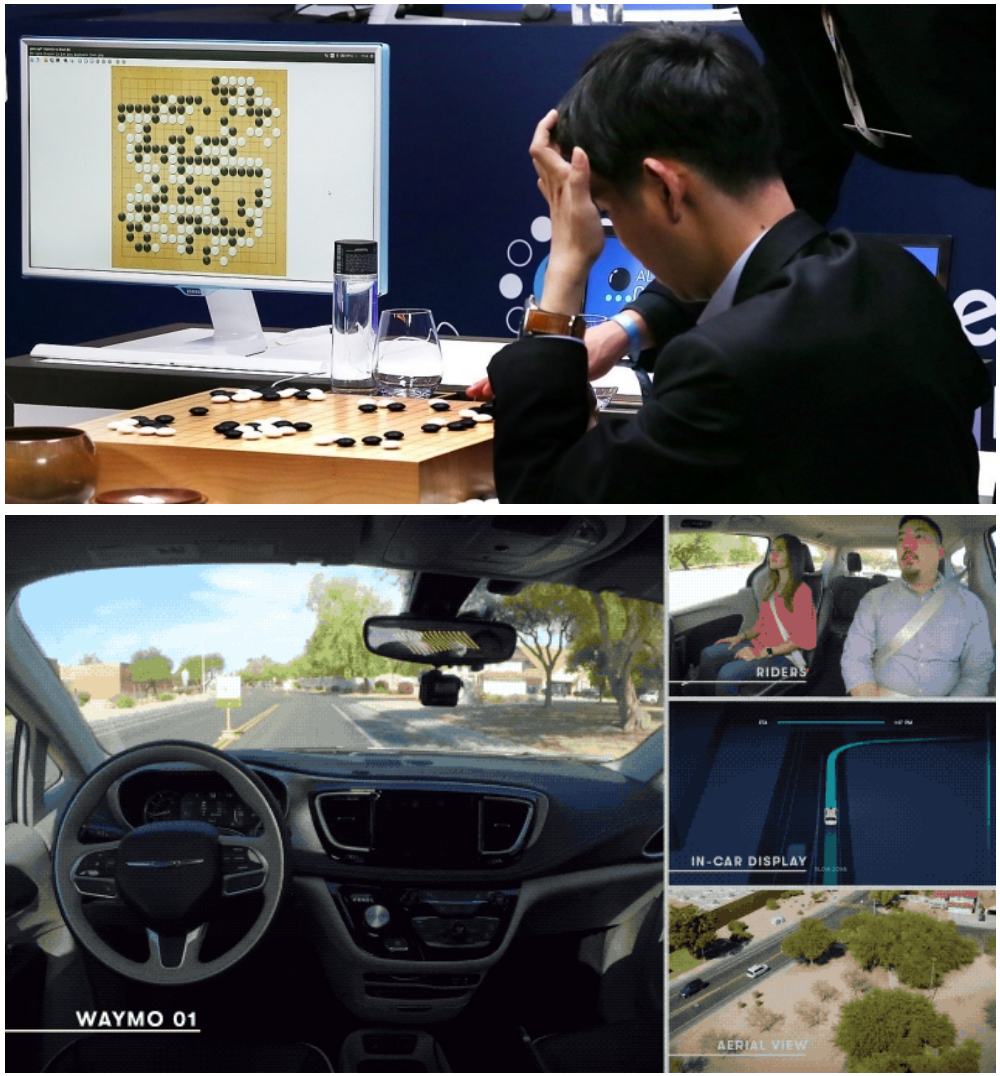
Today*



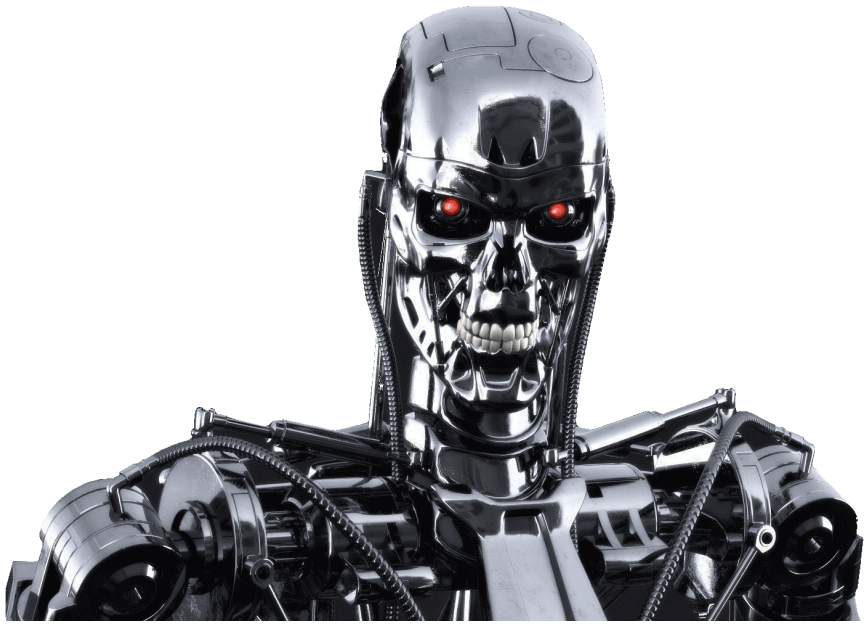
Towards generally intelligent agents?

- Artificial general intelligence
- AIXI
- Artifical life

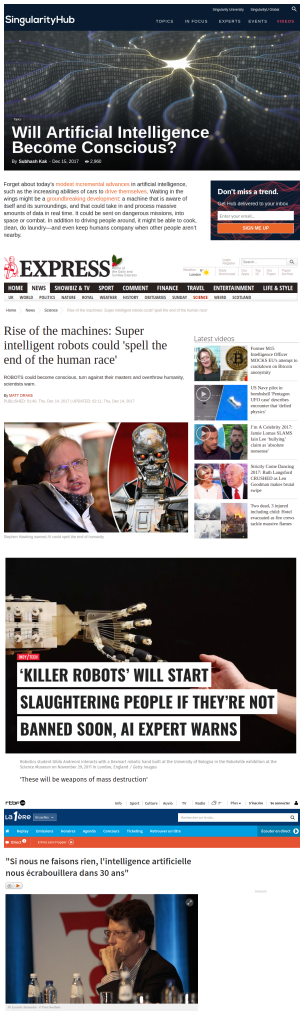
*: Take today's lecture with a grain of salt. Image credits: CS188, UC Berkeley.



From technological breakthroughs...



... to press coverage.



Artificial narrow intelligence

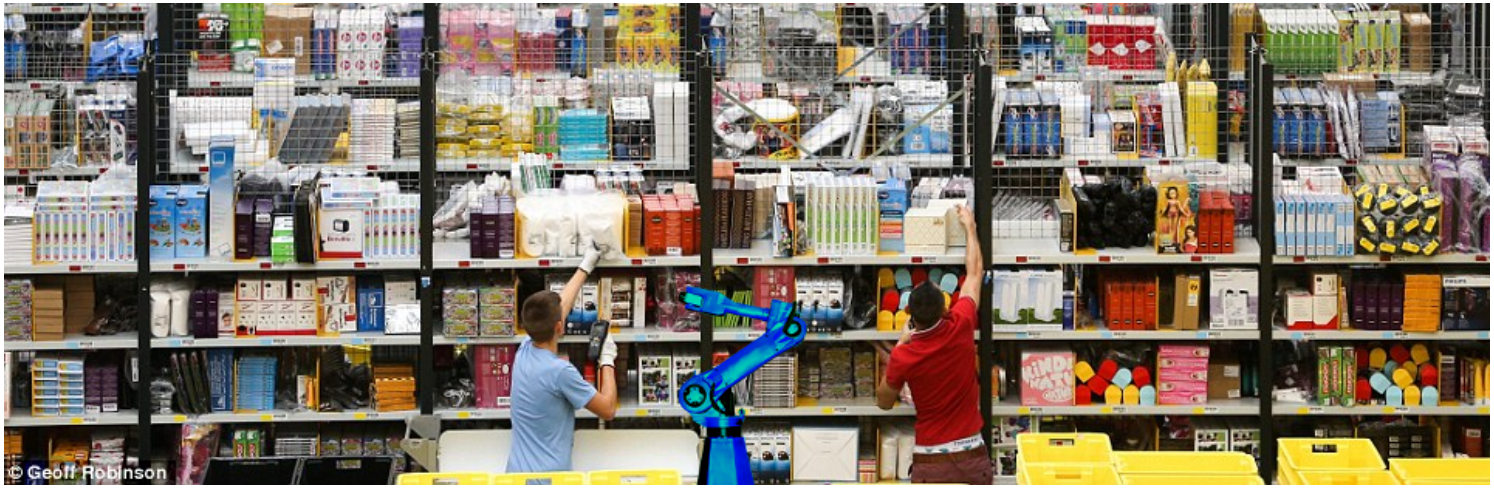
Artificial intelligence today remains narrow:

- 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.

The case of AlphaGo

Convenient properties of the game of Go:

- 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)



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

- Deterministic: Yes.
- Fully observed: Almost.
- Discrete action space: Yes
- Perfect simulator: Nope! Not at all.
- Short episodes: Not really...
- Clear and fast evaluation: Not good.
- Huge dataset available: Nope.

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 and definitive definition.
- Agreement that AGI is required to do the following:
  - reason, use strategy, solve puzzle, plan,
  - make judgments under uncertainty,
  - represent knowledge, including commonsense knowledge,
  - improve and learn new skills,
  - 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.

Irving John Good (1965)

Singularity

- 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.

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

Roads towards Artificial General Intelligence

Several working hypothesis:

- Supervised learning
- Unsupervised learning
- AIXI
- Artificial life
- Brain simulation

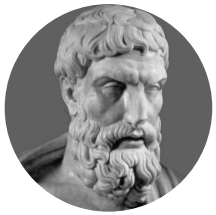
Or maybe (certainly) something else?

AIXI

In which environment is the agent?

- In general, we do not know!
- Solution:
  - maintain a prior over environments,
  - update it as evidence is collected,
  - follow the Bayes-optimal solution.

Occam: Prefer the simplest consistent hypothesis.



Epicurus: Keep all consitent hypotheses.

Bayes: $P(h|d) = \frac{P(d|h)P(h)}{P(d)}$

Turing: It is possible to invent a single machine which can be used to compute any computable sequence.

Solomonoff induction

- Use computer programs μ as hypotheses/environments.
- Make a weighted prediction based on all consistent programs, with short programs weighted higher.

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

- $\Upsilon(\pi)$ 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.
- $K(.)$ 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.

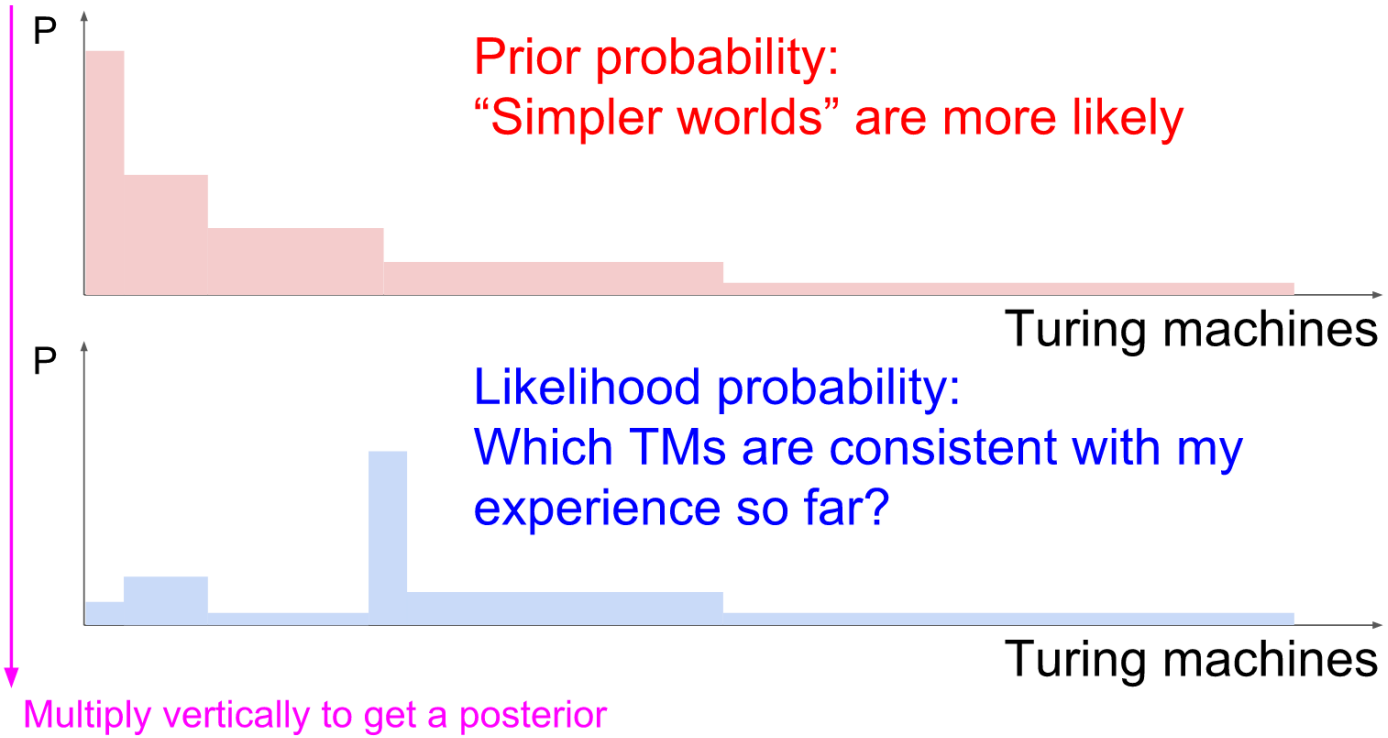
AIXI

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

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.



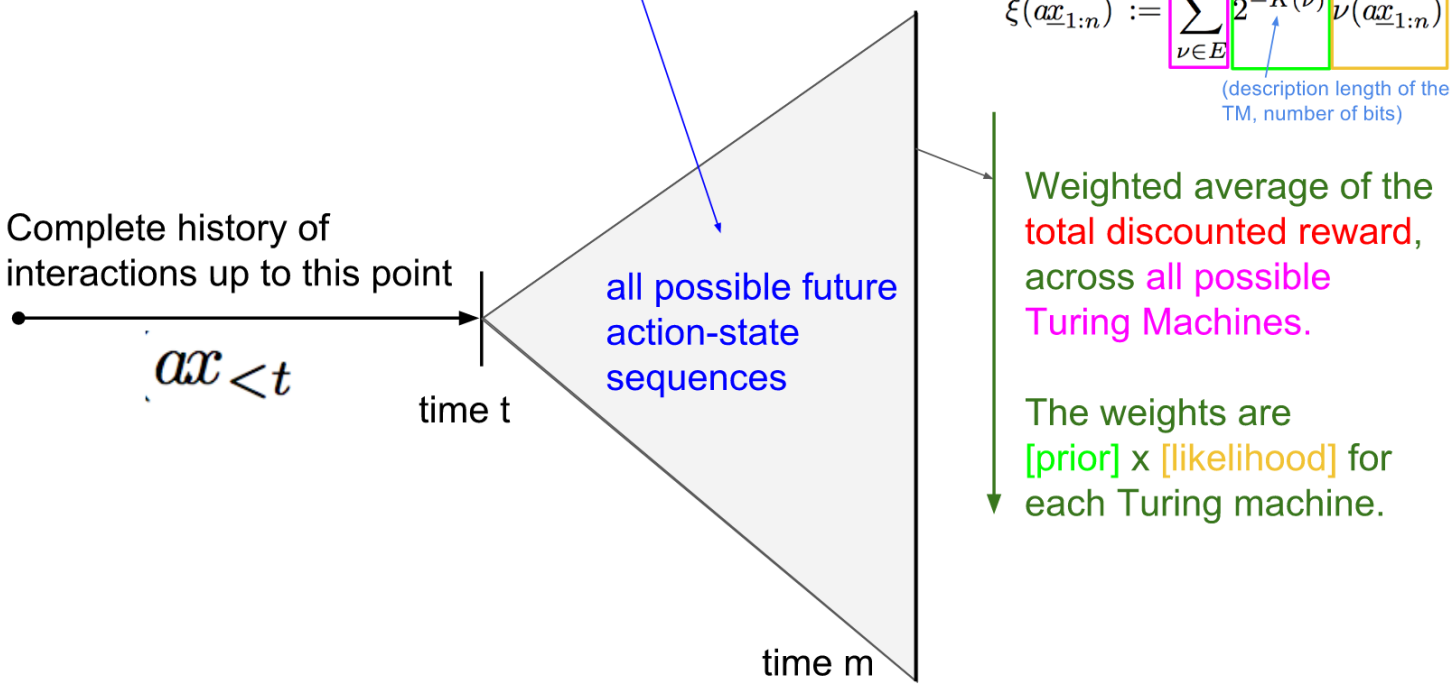
Acting optimally

- 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.

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

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

(description length of the TM, number of bits)



Incomputability

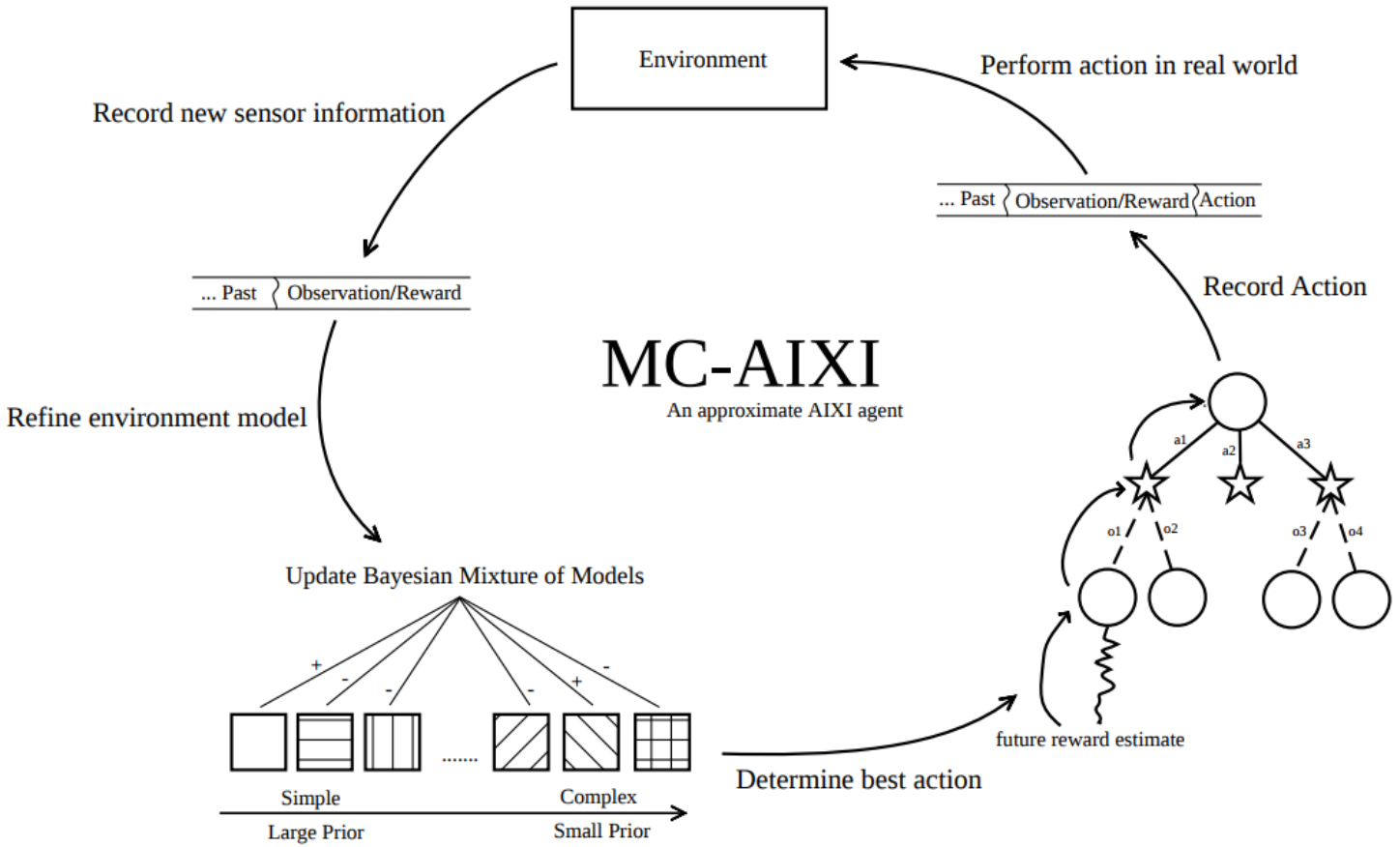
!!! !!!

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

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

Credits: Andrej Karpathy, Where will AGI come from?

Monte Carlo approximation

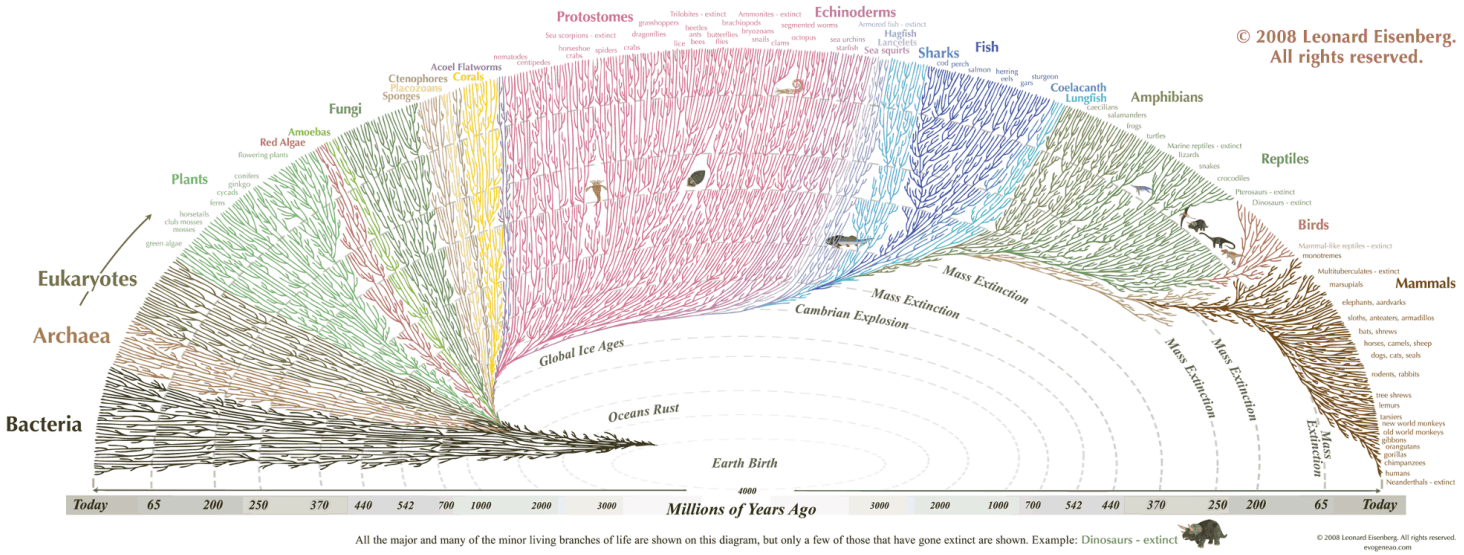


Benefits of a foundational theory of AI

AIXI provides

- high-level blue-print or inspiration for design;
- common terminology and goal formulation;
- understand and predict behavior of yet-to-be-built agents;
- appreciation of fundamental challenges (e.g., exploration-exploitation);
- definition/measure of intelligence.

Artificial life



How did intelligence arise in Nature?

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.

Wet artificial life: The line between life and not-life (Martin Hanczyc).

Evolution

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?
- 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.

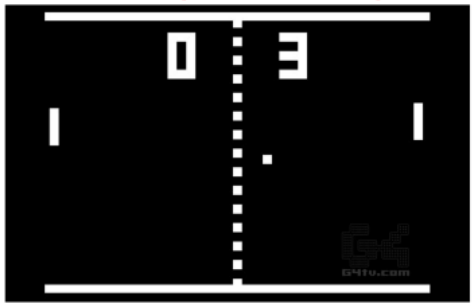

C

Neurevolution demo.

Environments for AGI?

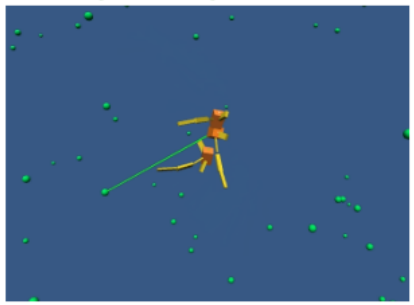
For the emergence of generally intelligent creatures, we presumably need environments that incentivize the emergence of a cognitive toolkit (attention, memory, knowledge representation, reasoning, emotions, forward simulation, skill acquisition, ...).

Doing it wrong:



Incentives a lookup table of correct moves.

Doing it right:



Incentivises cognitive tools.

Credits: Andrej Karpathy, Where will AGI come from?

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

Roads towards AGI

In order of (subjective) promisingness

- Artificial life
- Something not on our radar
- Supervised learning
- Unsupervised learning
- AIXI
- Brain simulation

What do you think?

A note of optimism: Don't fear intelligent machines,
work with them (Garry Kasparov).

Outline

- Lecture 1: Foundations
- Lecture 2: Solving problems by searching
- Lecture 3: Constraint satisfaction problems
- Lecture 4: Adversarial search
- Lecture 5: Representing uncertain knowledge
- Lecture 6: Inference in Bayesian networks
- Lecture 7: Reasoning over time
- Lecture 8: Making decisions
- Lecture 9: Learning
- Lecture 10: Communication
- Lecture 11: Artificial General Intelligence and beyond

Going further

This course is designed as an introduction to the many other courses available at ULiège and related to AI, including:

- ELEN0062: Introduction to Machine Learning
- INFO8004: Advanced Machine Learning
- INFO8010: Deep Learning
- INFO8003: Optimal decision making for complex problems
- INFO0948: Introduction to Intelligent Robotics
- INFO0049: Knowledge representation
- ELEN0016: Computer vision
- DROI8031: Introduction to the law of robots

Research opportunities

Feel free to contact us

- for research Summer internship opportunities (locally or abroad),
- MSc thesis opportunities,
- PhD thesis opportunities.

Thanks for following Introduction to Artificial Intelligence!

References

- 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.