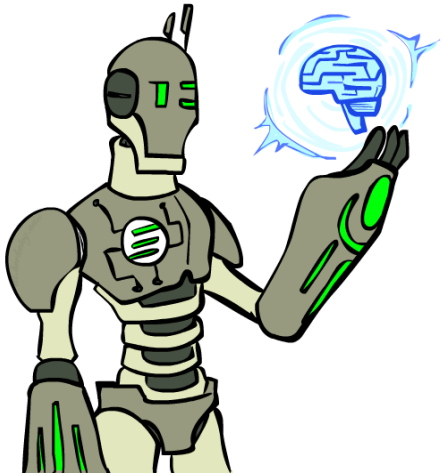
Introduction to Artificial Intelligence

Lecture 11: Artificial General Intelligence

Prof. Gilles Louppe
g.louppe@uliege.be

LIÈGE université

Today*



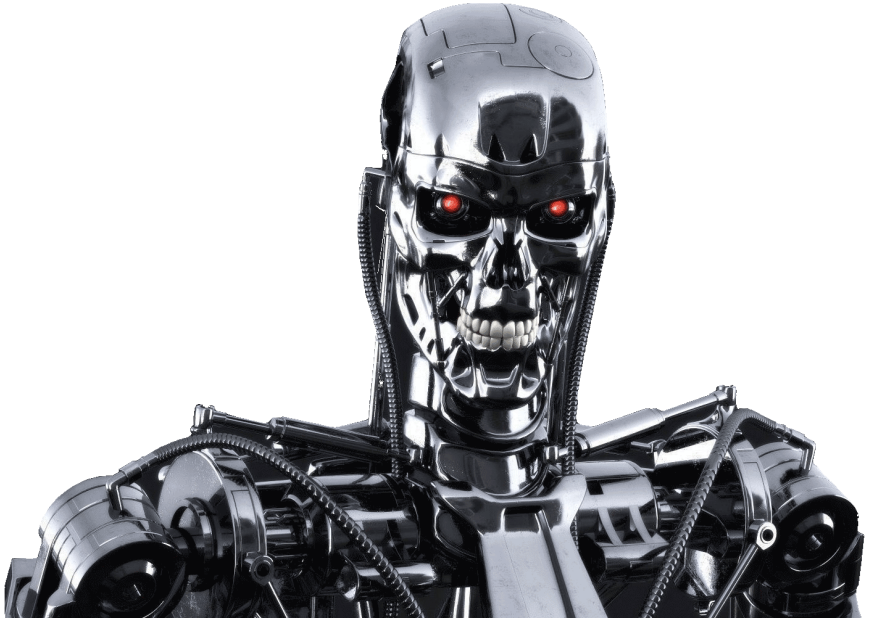
Towards generally intelligent agents?

- Artificial general intelligence
- AIXI
- Artifical life

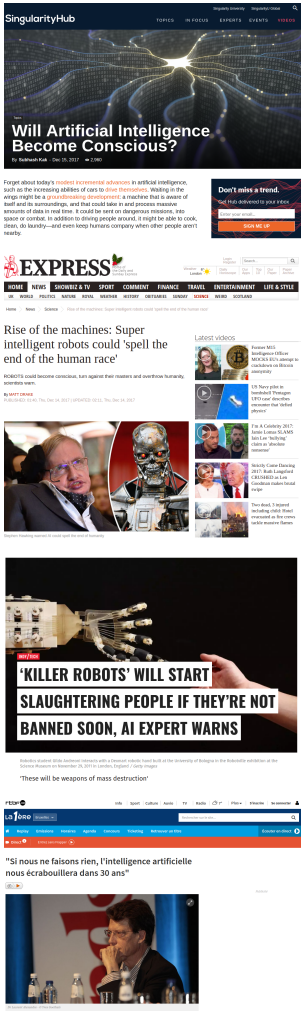
---

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

From technological breakthroughs...



... to press coverage.



Artificial narrow intelligence

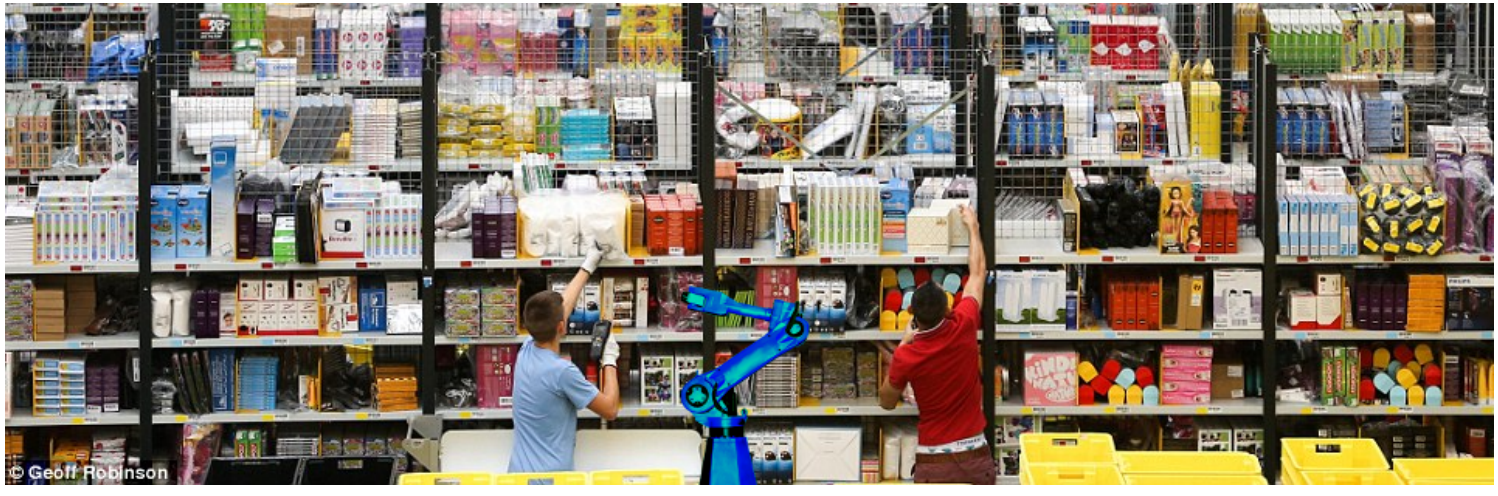
Today's artificial intelligence remains narrow:

- 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?

AGI

Artificial general intelligence, or AGI, is the intelligence of a machine that could successfully perform any intellectual task that a human being can perform.

The scientific community agrees that AGI would be 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.

Roads towards AGI

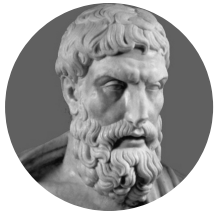
Several working hypothesis:

- Learning (supervised, unsupervised, reinforcement)
- AIXI
- Artificial life
- ... or probably something else?

AIXI

AIXI (Hutter, 2005) is a theoretical mathematical formalism of artificial general intelligence.

Occam: Prefer the simplest consistent hypothesis.



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

AIXI defines a measure of universal intelligence as

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

where

- $\Upsilon(\pi)$ 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.

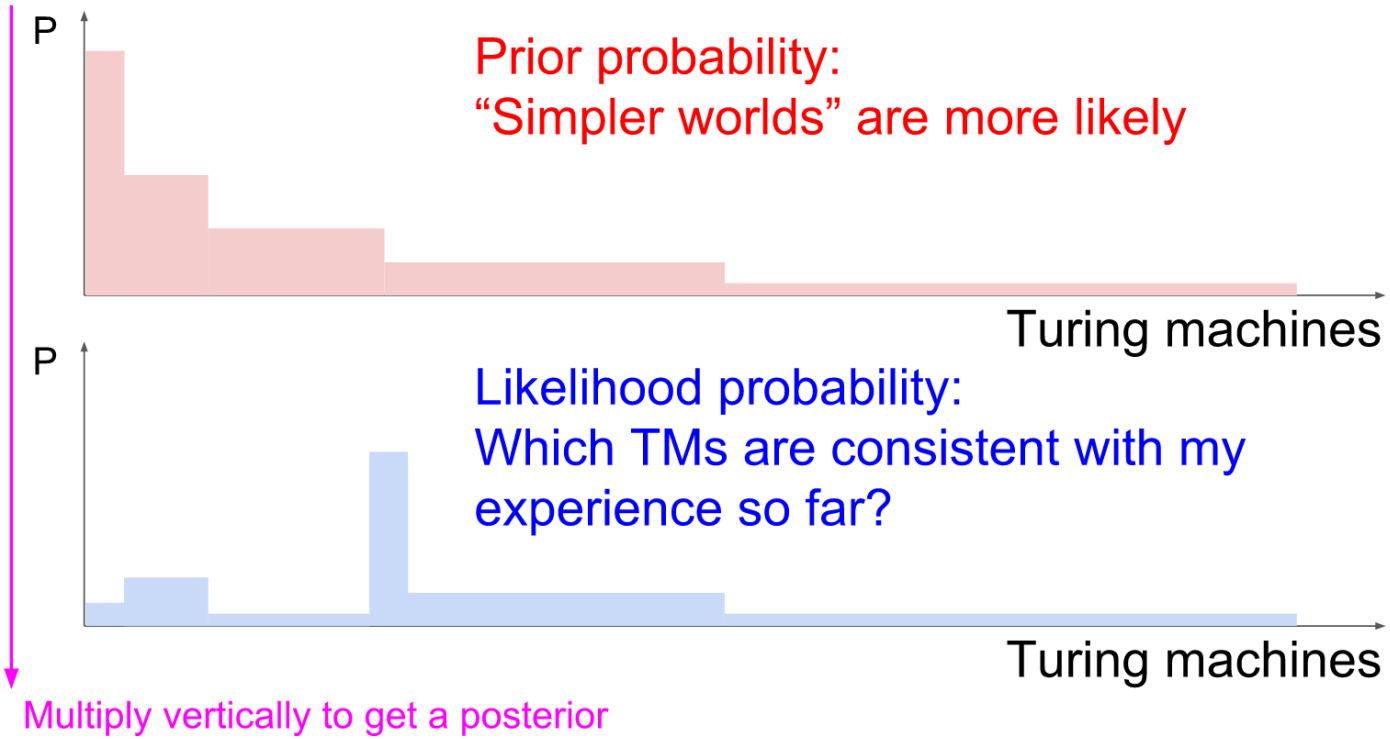
AIXI

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

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.



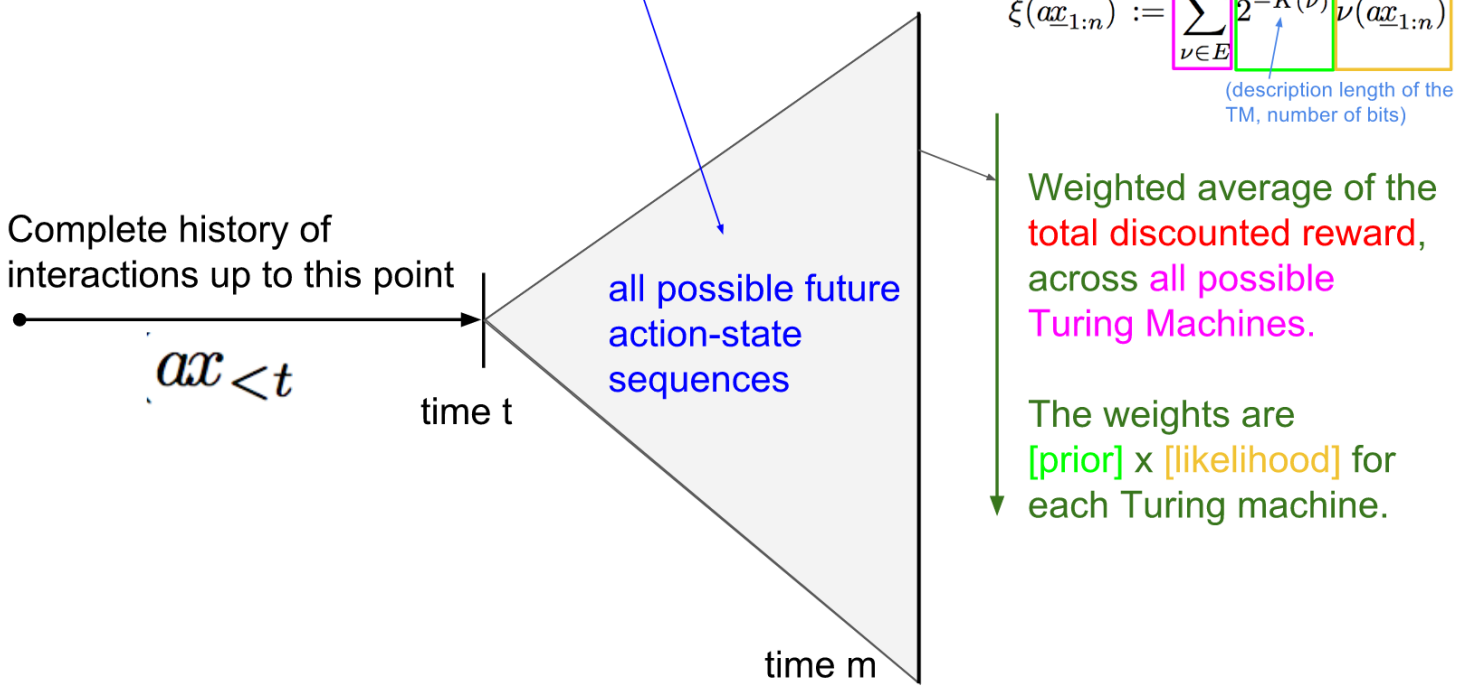
Acting optimally

- 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)



Credits: Andrej Karpathy, Where will AGI come from?

AIXI is incomputable

!!! !!!

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

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

—

Credits: Andrej Karpathy, Where will AGI come from?

Benefits of AIXI

The AIXI theoretical formalism of AGI provides

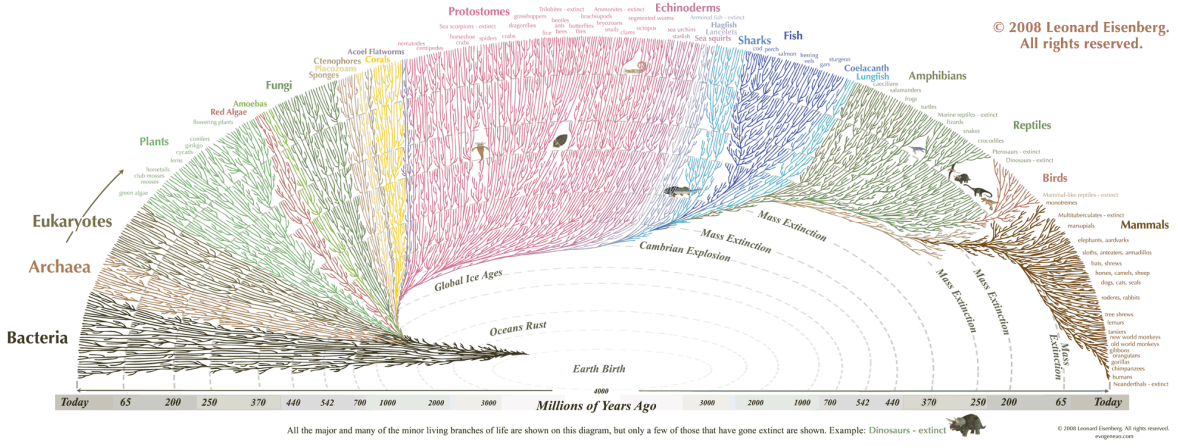
- a high-level blue-print or inspiration for design;
- common terminology and goal formulation;
- understand and predict behavior of yet-to-be-built agents;
- appreciation of fundamental challenges (e.g., exploration-exploitation);
- definition/measure of intelligence.

Artificial life

Artificial life

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.



How did intelligence arise in Nature?

Approaches

There are three main kinds of artificial life, named after their approaches:

- Software approaches (soft)
- Hardware approaches (hard)
- Biochemistry approaches (wet)

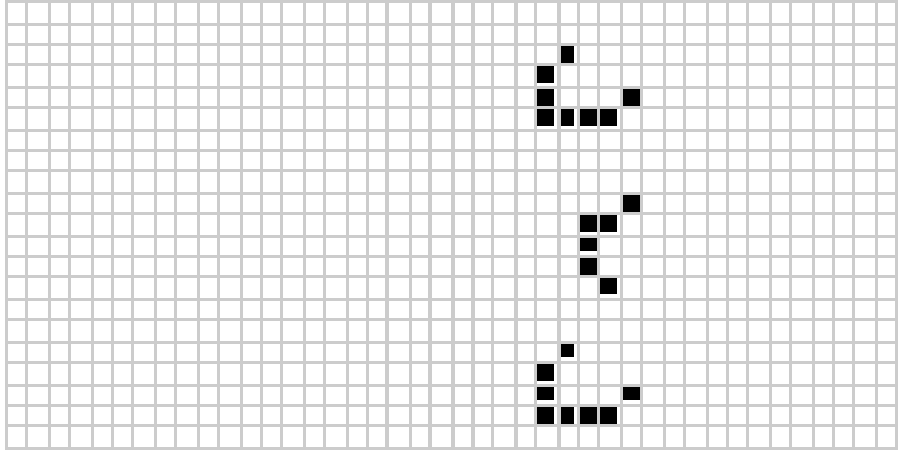
The field of AI has traditionally used a top down approach. Artificial life generally works from the bottom up.

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

Evolutionary algorithms

Evolution may hypothetically be interpreted as an (unknown) algorithm.

- This algorithm gave rise to AGI (e.g., it induced humans).
- Simulation of the evolutionary process should eventually reproduce life and, maybe, intelligence?



Conway's game of life

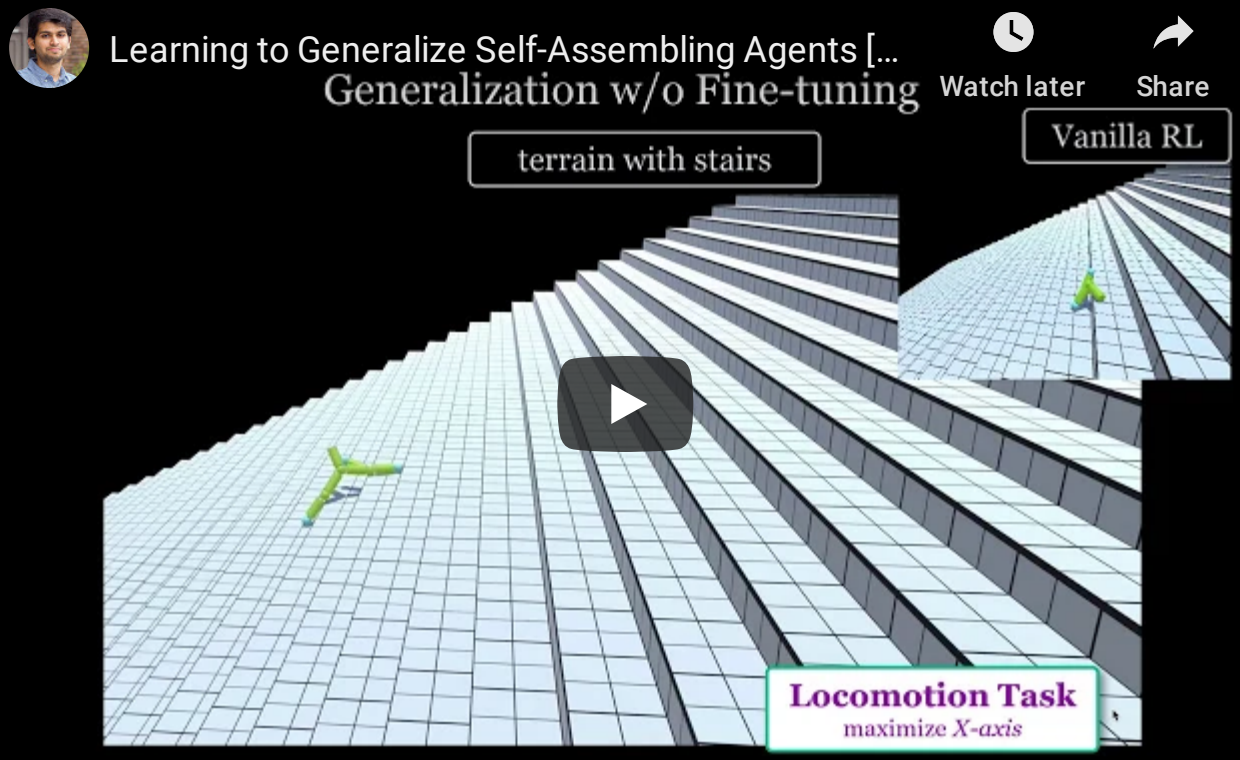
Conway's game of life

Evolutionary algorithms as metaheuristic optimization algorithms

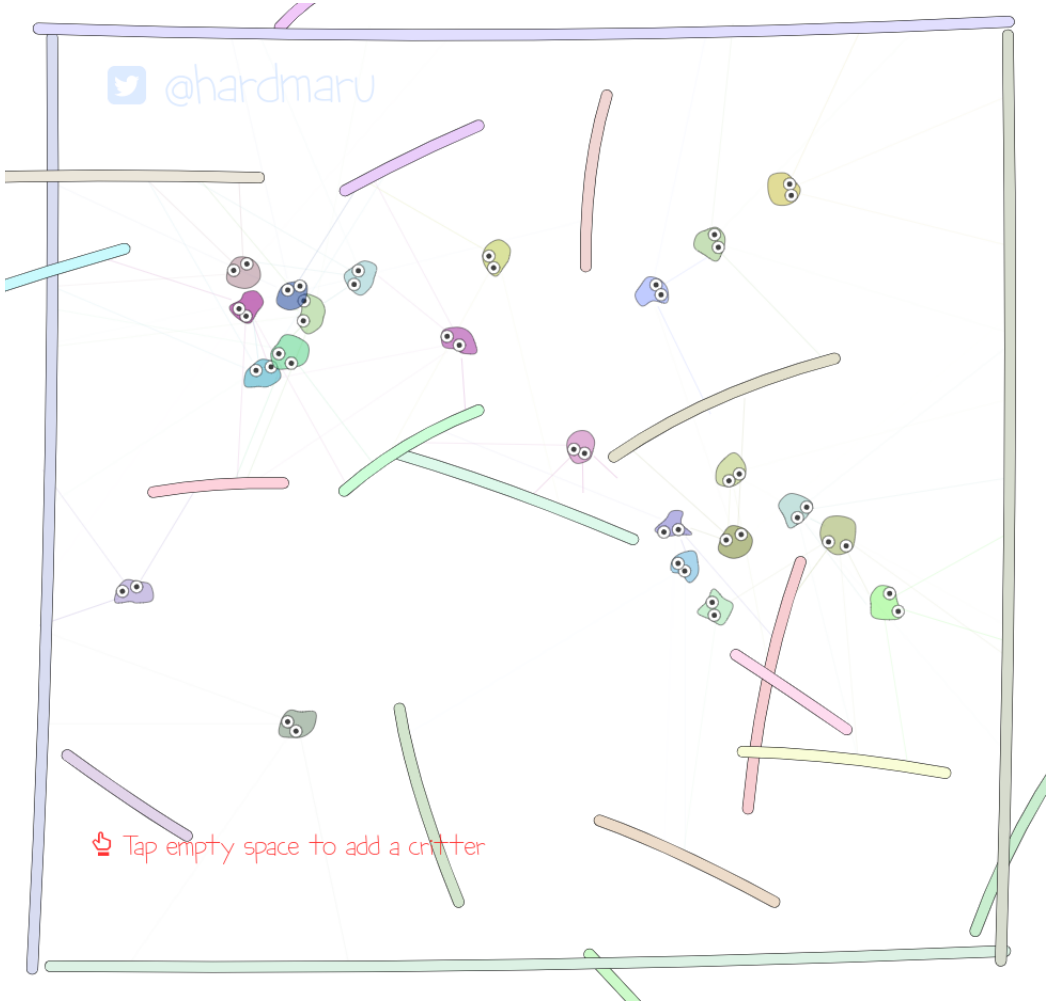
1. Start with a random population of creatures.
2. Repeat until termination:
    1. Each creature is tested for their ability to perform a given task.
    2. Select the fittest creatures for reproduction.
    3. Breed new creatures by combining and mutating the virtual genes of their selected parents.
    4. Replace the least-fit creatures of the population with new creatures.

As this cycle of variation and selection continues, creatures with more and more successful behaviors may emerge.

Karl Sims, 1994.



Self-assembling morphologies (Pathak et al, 2019)

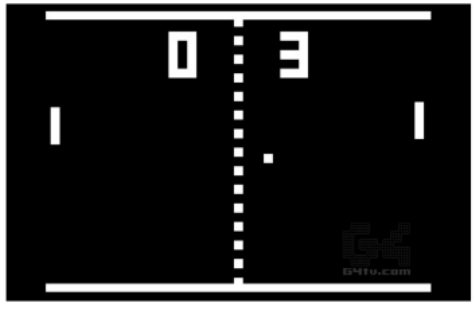


Creatures avoiding planks [demo].

Environments for AGI?

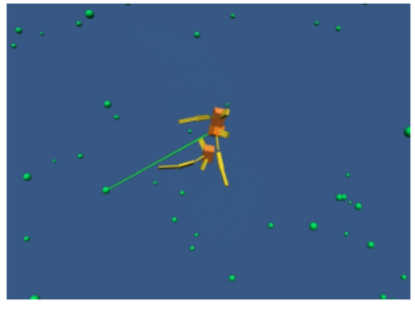
For the emergence of generally intelligent creatures, environments should incentivize the emergence of a cognitive toolkit (attention, memory, knowledge representation, reasoning, emotions, forward simulation, skill acquisition, ...).

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:

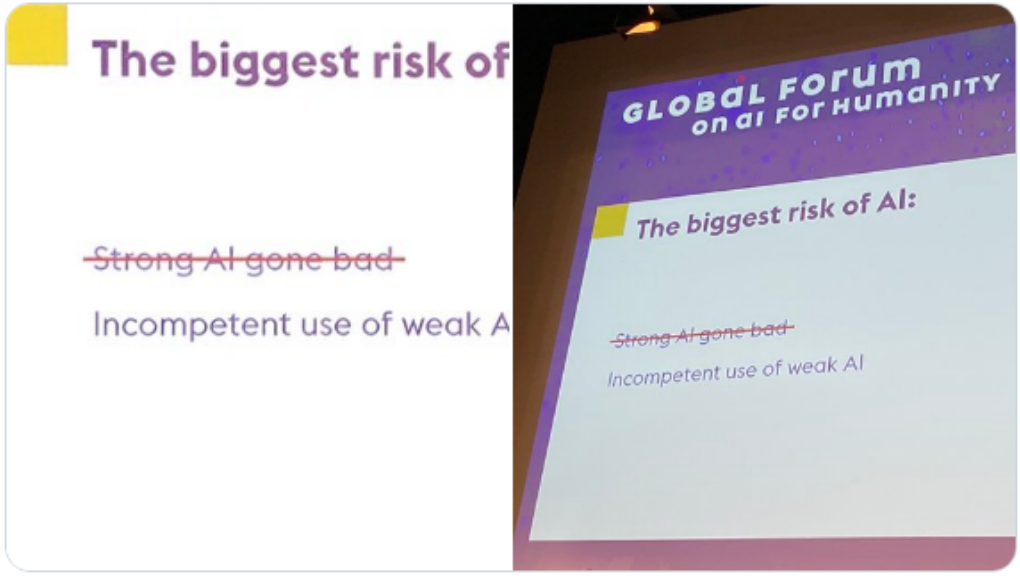
- Variety: the environment is parameterized by its agent population. The optimal strategy must be derived dynamically.
- Natural curriculum: the difficulty of the environment is determined by the skill of the other agents.

Conclusions

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

strategies.ai
@strategies_ai

Prof. @HolgerHoos (@UniLeidenNews) « Is the biggest danger a strong AI going bad? No. It's incompetent use of weak AI. ». Couldn't agree more. Right, @dekai123 @SachaAlanoca @NicolasMoes? 😉
#AI #AGI #hottopic #debate #GFAIH



Beyond Pacman

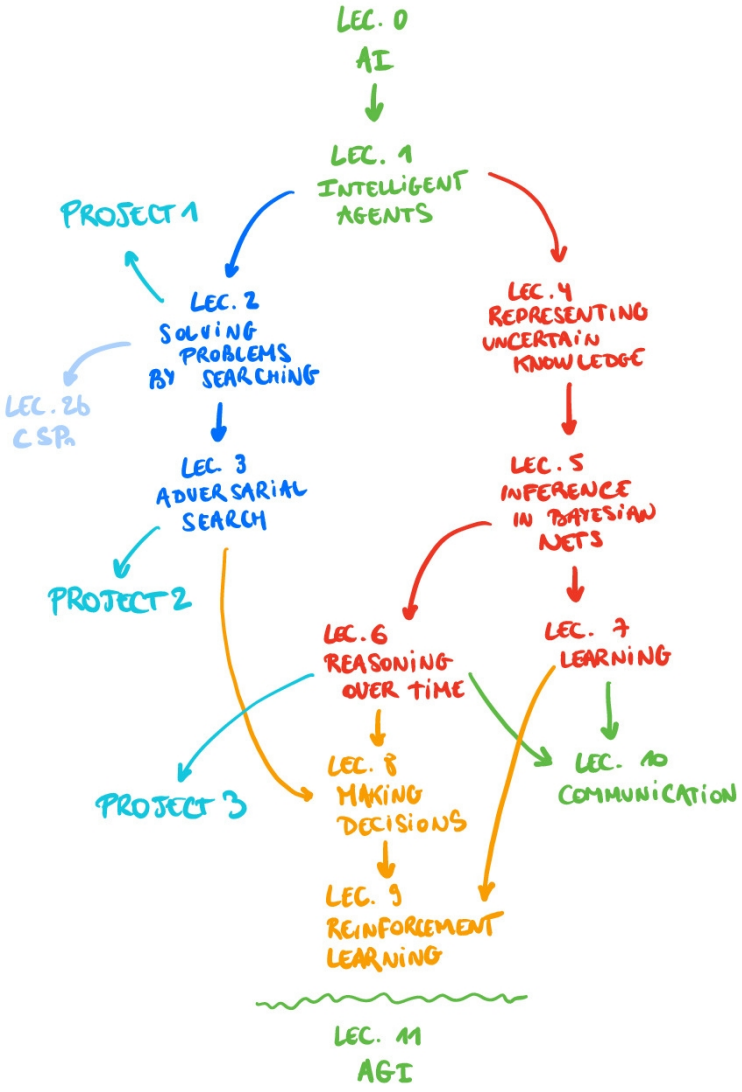
Artificial intelligence algorithms are transforming science, engineering and society.

As future engineers or scientists, AI offers you opportunities to address some of the world's biggest challenges. Seize them!

—
E.g.: Climate Change AI.

Recap

- Lecture 0: Artificial intelligence
- Lecture 1: Intelligent agents
- Lecture 2: Solving problems by searching
- Lecture 2b: Constraint satisfaction problems (optional)
- Lecture 3: Adversarial search
- Lecture 4: Representing uncertain knowledge
- Lecture 5: Inference in Bayesian networks
- Lecture 6: Reasoning over time
- Lecture 7: Learning
- Lecture 8: Making decisions
- Lecture 9: Reinforcement learning
- Lecture 10: Communication
- Lecture 11: Artificial General Intelligence and beyond


LEC. 0
AI
LEC. 1
INTELLIGENT AGENTS
PROJECT 1
LEC. 2
SOLVING PROBLEMS BY SEARCHING
LEC. 26
CSP
LEC. 3
ADVERSARIAL SEARCH
PROJECT 2
LEC. 4
REPRESENTING UNCERTAIN KNOWLEDGE
LEC. 5
INFERENCE IN BAYESIAN NETS
LEC. 6
REASONING OVER TIME
LEC. 7
LEARNING
PROJECT 3
LEC. 8
MAKING DECISIONS
LEC. 10
COMMUNICATION
LEC. 9
REINFORCEMENT LEARNING
LEC. 11
AGI

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
- ELEN0060: Information and coding theory
- MATH2022: Large-sample analysis: theory and practice
- DROI8031: Introduction to the law of robots

Research opportunities

Feel free to contact us

- for research Summer internship opportunities
- MSc thesis opportunities
- PhD thesis opportunities

Thanks for following Introduction to Artificial Intelligence!