Elements of DSAI

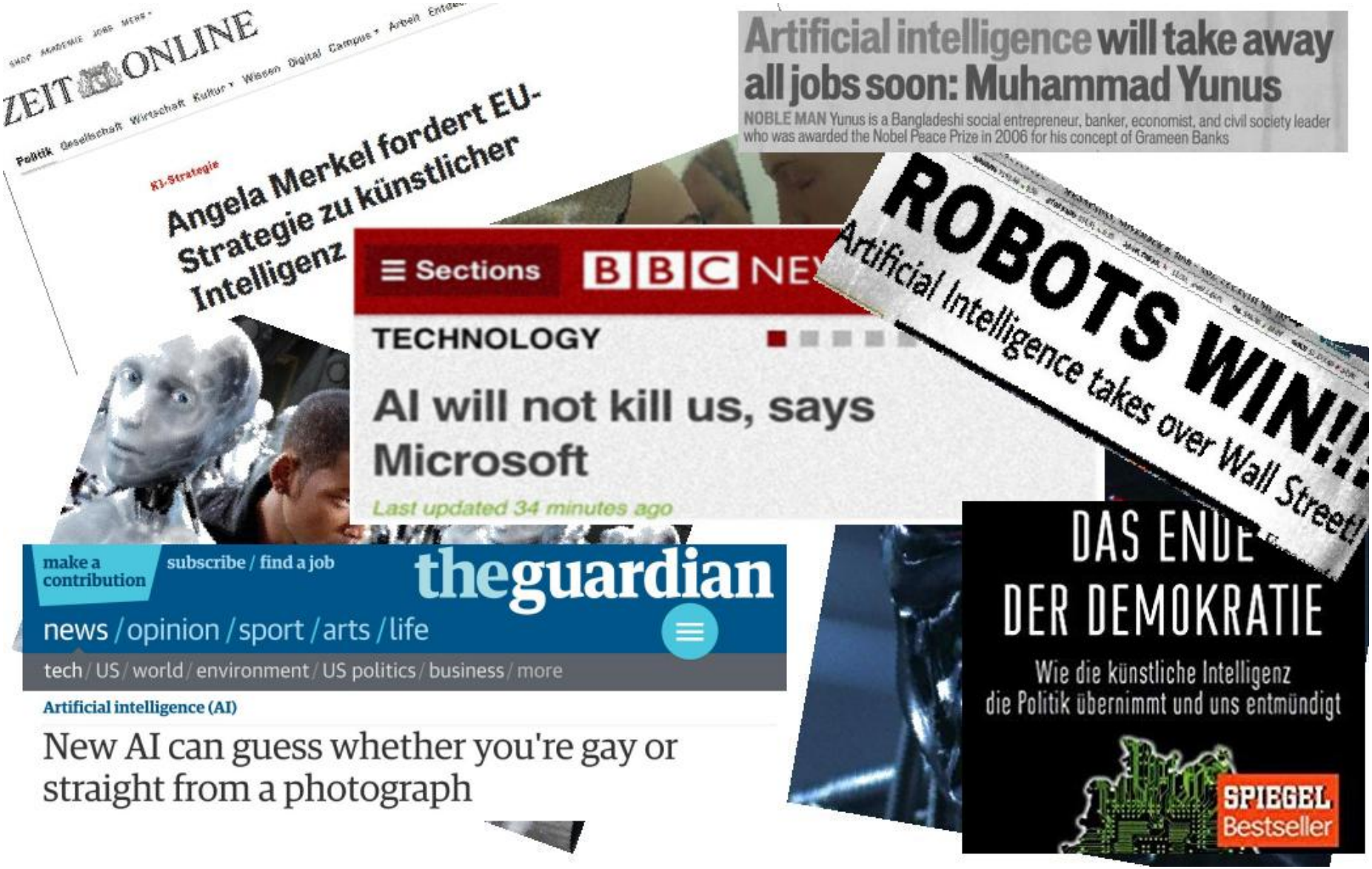
AI History and Concepts

What is AI?

Jörg Hoffmann

SAARLAND UNIVERSITY
COMPUTER SCIENCE

Winter Term 2019/20


ZEIT ONLINE
Angela Merkel fordert EU-Strategie zu künstlicher Intelligenz
Artificial intelligence will take away all jobs soon: Muhammad Yunus
NOBLE MAN Yunus is a Bangladeshi social entrepreneur, banker, economist, and civil society leader who was awarded the Nobel Peace Prize in 2006 for his concept of Grameen Banks
ROBOTS WIN!!!
Artificial Intelligence takes over Wall Street!
Sections
TECHNOLOGY
AI will not kill us, says Microsoft
Last updated 34 minutes ago
make a contribution
subscribe / find a job
theguardian
news / opinion / sport / arts / life
tech / US / world / environment / US politics / business / more
Artificial intelligence (AI)
New AI can guess whether you're gay or straight from a photograph
DAS ENDE DER DEMOKRATIE
Wie die künstliche Intelligenz die Politik übernimmt und uns entmündigt
SPIEGEL Bestseller

Agenda: AI Research History and Concepts

- **AI?** What does this term even mean?
  → Spoiler: Nobody knows.
- **The Research Field AI:** "AI" is, foremost, a name.
  → History and current composition.
- **Rationality:** What does it mean to be rational?
  → We'll discuss the common stand-in for "intelligence".
- **Agents & Environments:** In AI ...
  → Some basic concepts and terminology.

What is Intelligence?

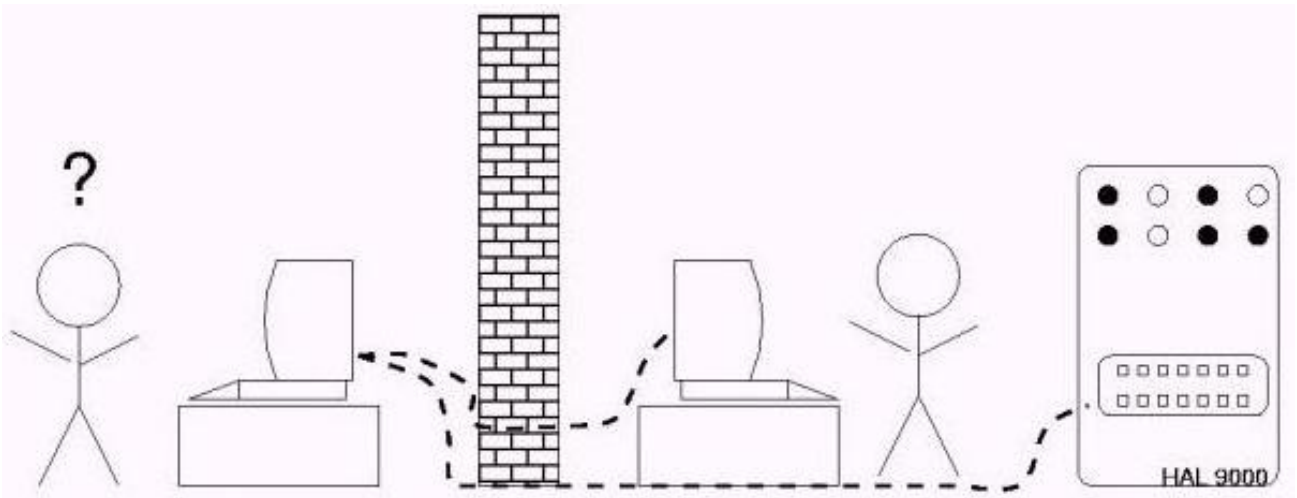
Nobody knows:

- Ability to think? (what does this mean?)
- Ability to learn? (what does this mean?)
- Creativity? (what does this mean?)
- Being good at maths? (what else?)
- Being good at Chess or Go? (what else?)
- Passing an IQ test with high marks? (go away!)

$\rightarrow$ This question has been debated in Philosophy since centuries ...

What is *Artificial* Intelligence?

Take 1: The Turing Test



- Suggested (in various forms) by Alan Turing as a measurable definition of AI.
- Yearly competitions: Loebner Prize
- Recent serial winner: Mitsuku

What is *Artificial* Intelligence?

Take 2: Let's try to be systematic here ...

	Humanly	**Rationally**
Thinking	Cognitive Science Neural Networks? Certainly not yet!	Logics Machine Learning
Acting	Turing Test	APPLICATIONS

→ Note: Thinking is (sometimes?) a prerequisite for acting ... logics and machine learning are motivated by, and very useful in, applications!

The Four Categories: Summary

Acting Humanly: Turing Test. Not much pursued otherwise.

≈ Aeronautics: "Machines that fly so exactly like pigeons that they can even fool other pigeons".

Thinking Humanly: Cognitive Science. How do humans think, how does the human brain work.

→ Neural networks are an (extremely simple, so far) approximation.

Thinking Rationally: Logics (formalization of knowledge and deduction). Machine Learning (ML) (mathematical formulation of learning).

Acting Rationally: How to make good action choices?

→ Is what we're interested in, in practice. Encompasses logics and ML (in particular neural networks) as methods to take rational decisions.

The History of AI

Origins: The dream of an "artificial intelligence" (broadly interpreted) is age-old (Philosophy mainly).

1956: Inception of AI at Dartmouth Workshop. John McCarthy proposes the name "Artificial Intelligence". Early enthusiasm, famous quote:

"It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in the visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied."

60's: Early successes. "Intelligent Behavior" is shown in many demonstration systems for microworlds (Blocksworld).

70's: How to scale from microworlds to real applications?
→ Knowledge-based systems, knowledge provided by humans.

Early 80's: Commercial success of rule-based expert systems.

The History of AI, ctd.

Late 80's: Expert systems prove less promising than imagined (difficult to update/maintain, cannot learn, brittle). → "AI Winter".

90's–00's: Formalization of AI techniques and increased use of mathematics in the field. Quote from [Russell and Norvig (1995)]:

"A better understanding of the problems and their complexity properties, combined with increased mathematical sophistication, has led to workable research agendas and robust methods."

10's: Re-advent of neural networks (NN).

NN have decade-old roots. Almost forgotten in the 90's and 00's. "Sudden" success in image classification end of 00's, way better than human-coded rules. Since then, rapid successes and hype. "Deep" NN = several layers (how many? next question please).

Enablers: better algorithms for training; lots of data; hardware.

20's: Computers rule the world? Next AI Winter?

AI Sub-Areas

Search: How to effectively find solutions in problems with large search spaces (combinatorial explosion).

→ International Symposium on Combinatorial Search (SOCS)

KR: Knowledge representation and reasoning (logic and deduction).

→ International Conference on Principles of Knowledge Representation and Reasoning (KR)

SAT & CP: General formulation and solution of search problems that involve satisfying a set of constraints.

→ I.C. Theory and Applications of Satisfiability Testing (SAT), I.C. Principles and Practice of Constraint Programming (CP)

Planning: General formulation and solution of search problems that involve finding goal-leading action strategies.

→ I.C. Automated Planning and Scheduling (ICAPS)

Uncertainty: Reasoning about uncertain knowledge.

→ I.C. Uncertainty in Artificial Intelligence (UAI)

AI Sub-Areas, ctd.

ML: Machine Learning.

→ e.g. I.C. Machine Learning (ICML), Conference on Neural Information Processing Systems (NeurIPS)

Multi-Agents: How to control/analyze systems of agents perceiving/acting individually?

→ I.C. Autonomous Agents and Multiagent Systems (AAMAS)

Robotics: How to control/design robots?

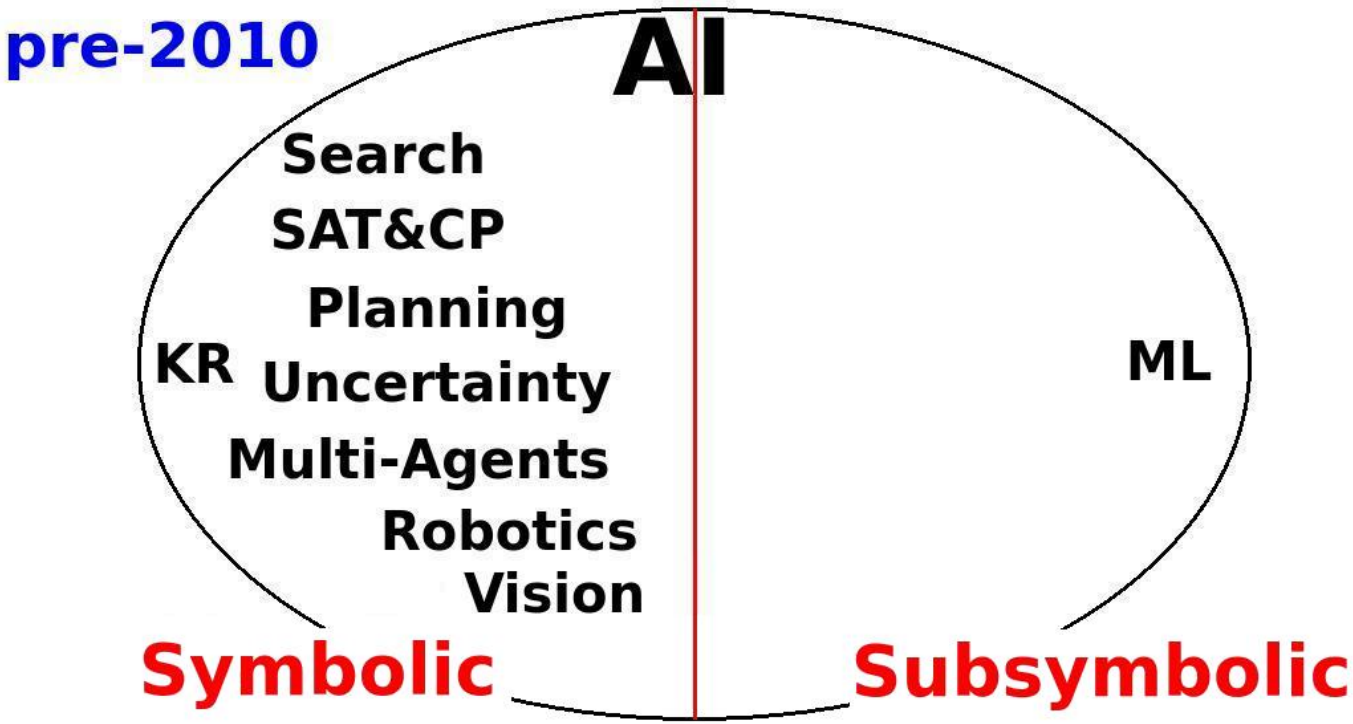
→ e.g. I.C. Robotics and Automation (ICRA), Robotics: Science and Systems (RSS)

Vision: How to interprete/analyze camera input?

→ e.g. Conference on Computer Vision and Pattern Recognition (CVPR), I.C. Computer Vision (ICCV)

→ Global, cross-AI, conferences: International Joint Conference on AI (IJCAI), American Conference on AI (AAAI), European Conference on AI (ECAI).

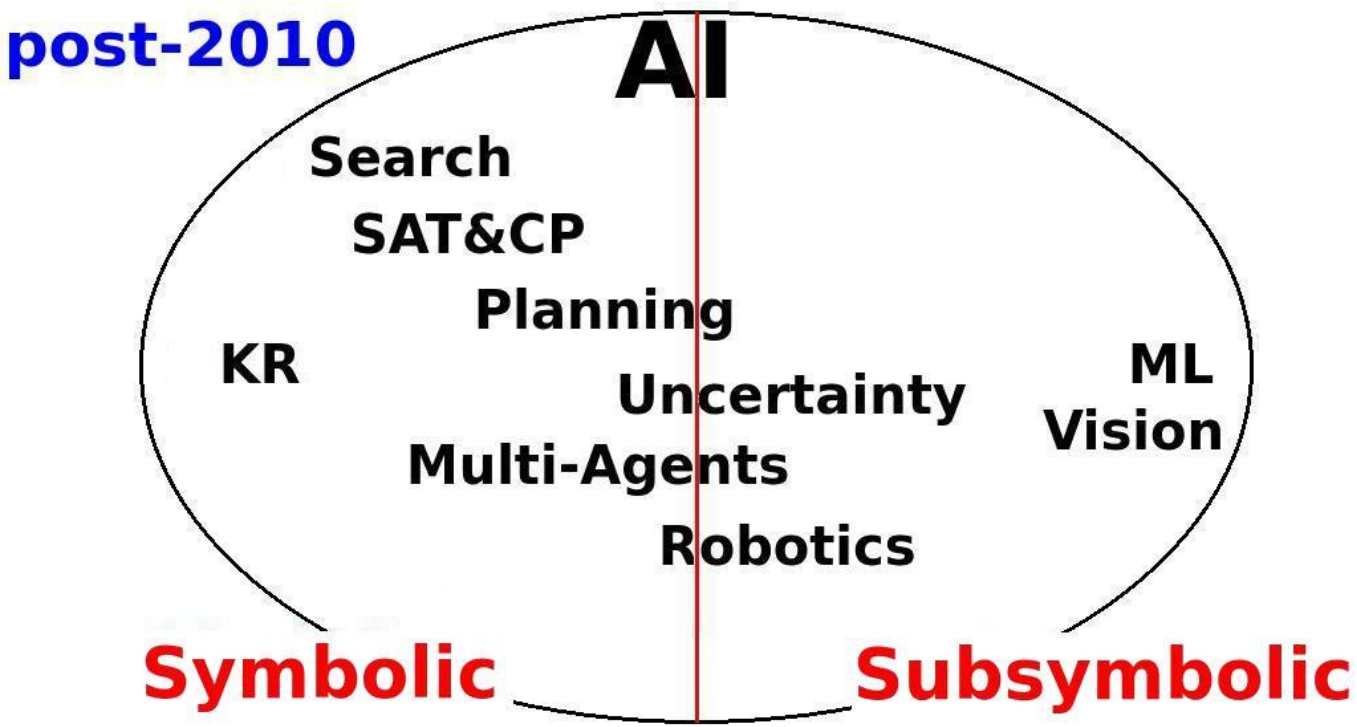
Symbolic vs. Subsymbolic AI



Symbolic: Conceptual (human-readable) formalization of world behavior

Subsymbolic: Mathematical fitting of data (world behavior observations)

Symbolic vs. Subsymbolic AI



Symbolic: Conceptual (human-readable) formalization of world behavior

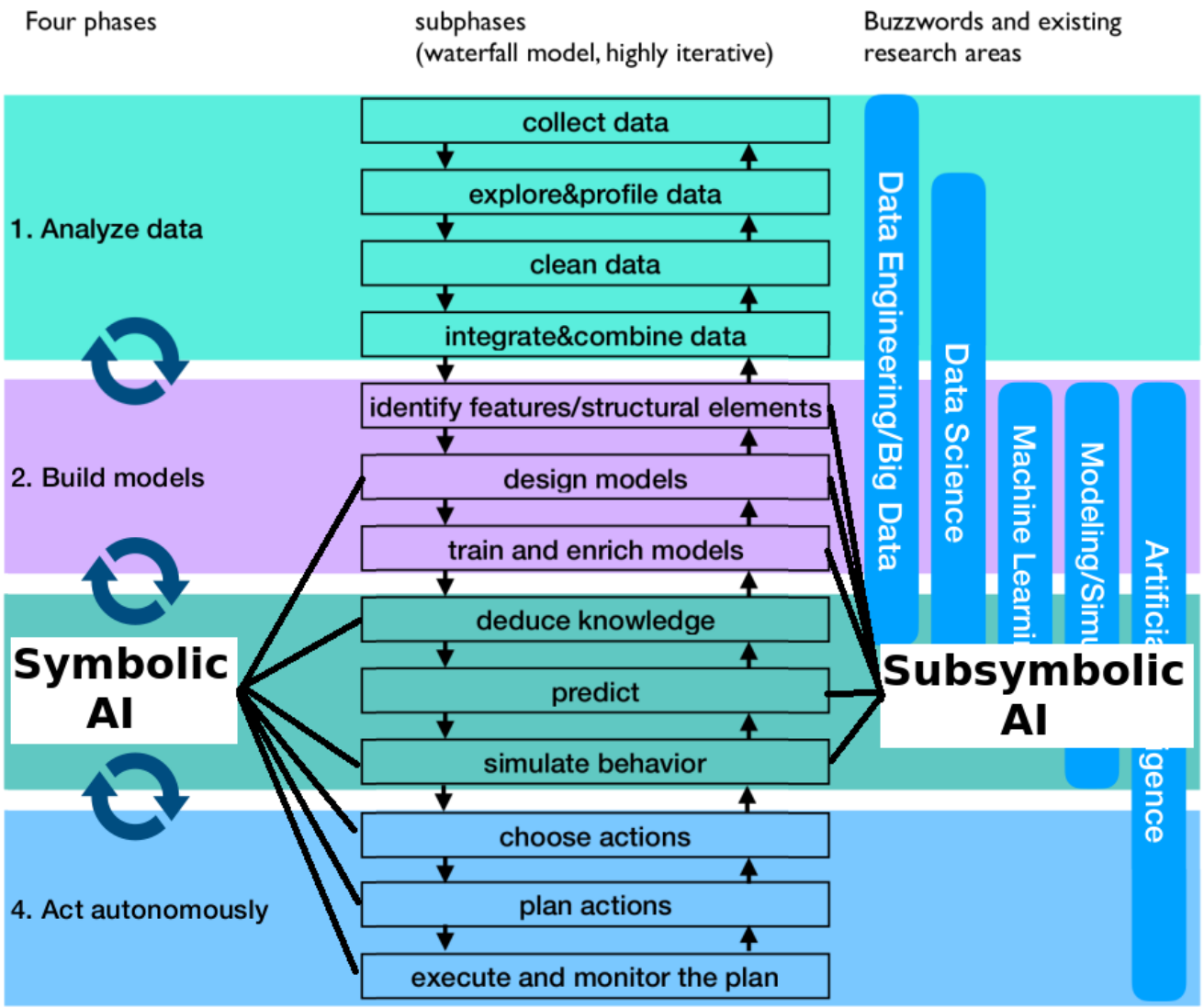
- Pros: instant performance, verifiability, explainability
- Cons: modeling can be costly or impossible, complexity of reasoning

Subsymbolic: Mathematical fitting of data (world behavior observations)

- Pros: highly performant, able to tackle problems elusive for conceptual modeling
- Cons: learning curve, opaque, hyperparameters difficult to set

→ **How to combine the two?**

AI and the DSAI Phase Model



Behaving Rationally . . .

. . . means to do "the right thing".

→ Select actions maximizing a performance measure.

Example: Performance measure of an autonomous vacuum cleaner

- m^2 per hour.
- Level of cleanliness.
- Energy usage.
- Noise level.
- Safety (behavior towards hamsters/small children).

Actually, Behaving Rationally . . .

. . . means to ATTEMPT to do "the right thing".

Examples:

- I check the traffic before crossing the street. As I cross, I am hit by a meteorite. Was I lacking rationality?
- Is there dirt under this bed? Is this a hamster or a shoe?

$\rightarrow$ The hypothetical best case ("the right thing") is often unattainable.

Map your input to the best possible output:

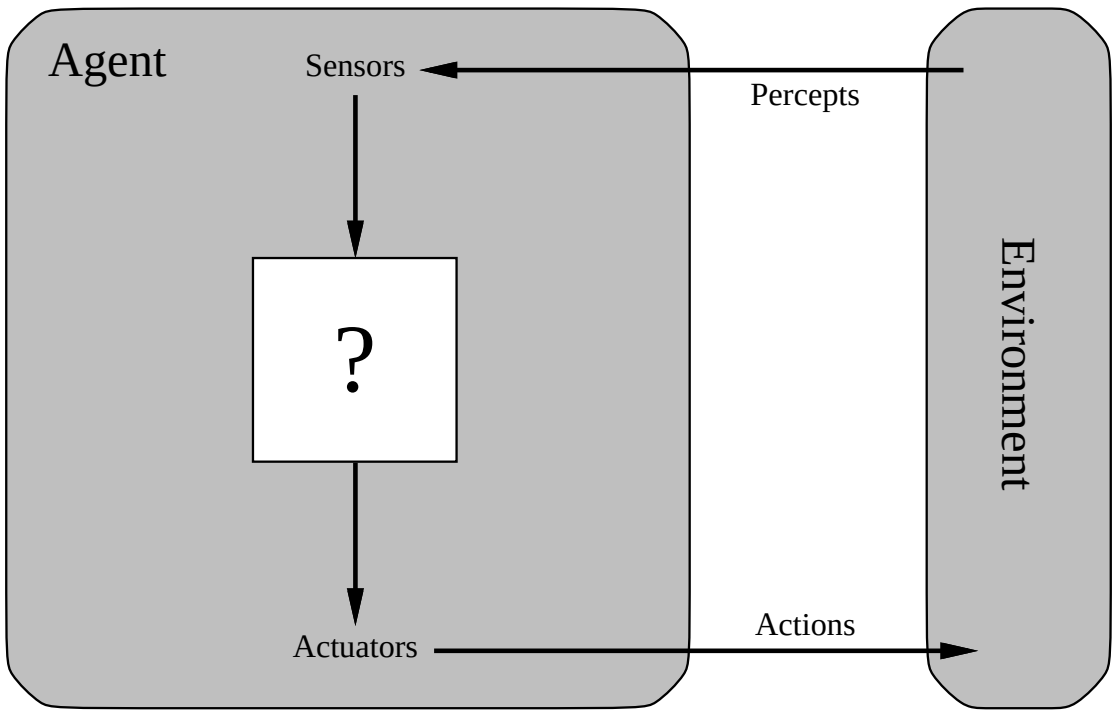
Performance measure M $\times$ Percepts P $\times$ Knowledge K $\rightarrow$ Action a

$\rightarrow$ a is optimal if it maximizes the expected value of M, given the evidence P and K. Rationality $=$ always choosing an optimal a.

What is an "Agent" in AI?

Agents:

- Perceive the environment through sensors ($\rightarrow$ percepts).
- Act upon the environment through actuators ($\rightarrow$ actions).



$\rightarrow$ Examples? Humans, animals, robots, software agents (softbots), ...

Table-Driven Agents

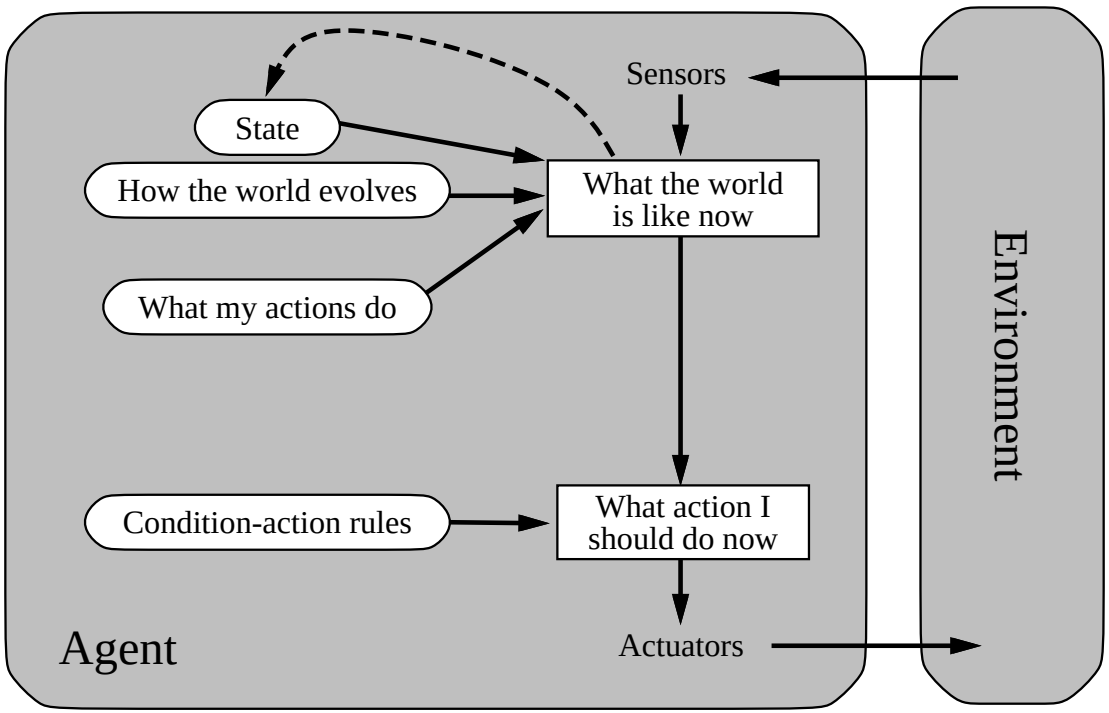
function TABLE-DRIVEN-AGENT(*percept*) **returns** an action
  persistent: *percepts*, a sequence, initially empty
    table, a table of actions, indexed by percept sequences, initially fully specified

  append *percept* to the end of *percepts*
  action ← LOOKUP(*percepts*, *table*)
  return *action*

→ Why is this not a good idea?

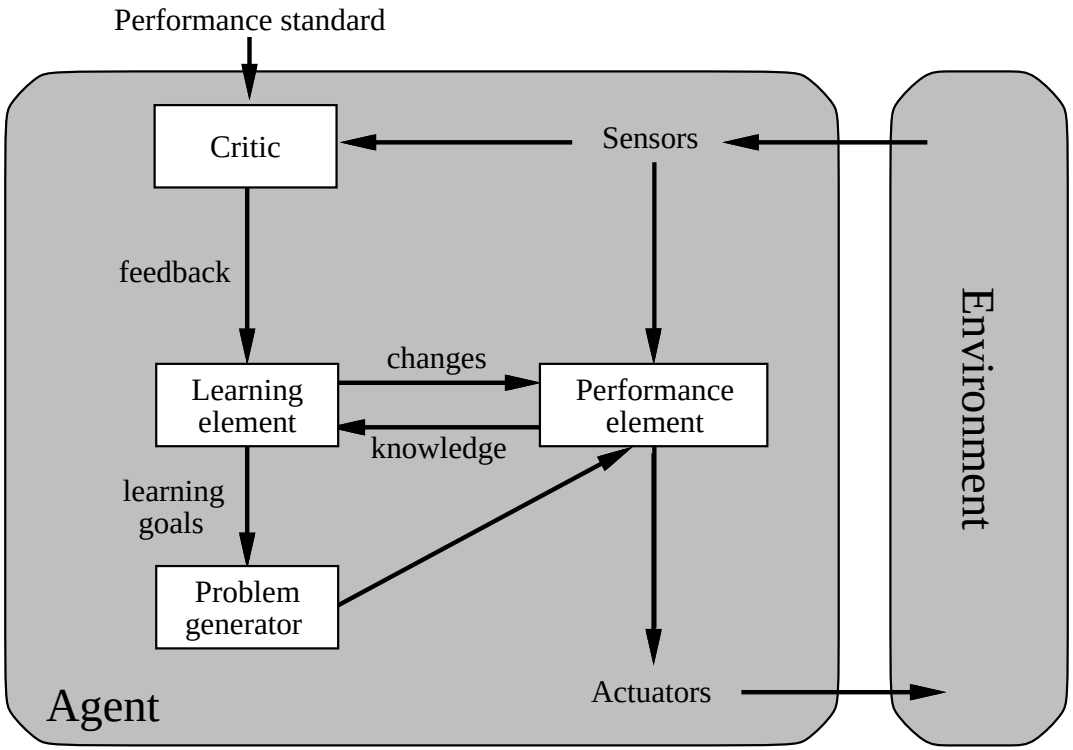
- The table is way too large. (Even when limiting to n binary percepts whose order of occurence does not matter, we have 2^n entries . . .)
- Who is supposed to write down this table anyway, even if it's got "only" a million entries?

Reflex Agents



→ Example? Vacuum cleaner: If it's dirty where you are right now, clean; otherwise, move somewhere else randomly.

Learning Agents



- Performance element: Selects actions (exploitation).
- Learning element: Learns new knowledge.
- Critic: Measures performance.
- Problem generator: Suggests actions favoring informative learning experiences (exploration).

Environments

- Accessible vs. inaccessible (fully observable vs. partially observable)

  Are the relevant aspects accessible to the sensors?

- Deterministic vs. stochastic

  Is the next state of the environment completely determined by the current state and the selected action?

- Episodic vs. sequential

  Can the quality of an action be evaluated within an episode (perception + action), or are future developments decisive?

- Static vs. dynamic

  Can the environment change while the agent is deliberating?

- Discrete vs. continuous

  Is the environment discrete or continuous?

- Single agent vs. multi-agent

  One agent or several? Competitive or collaborative?

Examples of Environments

Task	Observable	Deterministic	Episodic	Static	Discrete	Agents
Chess/Go without clock	fully	strategic	sequential	static	discrete	multi
Poker	partially	stochastic	sequential	static (?)	discrete	multi
Car driving	partially	stochastic	sequential	dynamic	continuous	multi
Medical diagnosis	partially	stochastic	episodic	dynamic	continuous	single
Image analysis	fully	deterministic	episodic	semi	continuous	single
Part-picking robot	partially	stochastic	episodic	dynamic	continuous	single
Refinery controller	partially	stochastic	sequential	dynamic	continuous	single
Interactive English tutor	partially	stochastic	sequential	dynamic	discrete	multi

$\rightarrow$ These properties may depend on the design: E.g., if the medical diagnosis system interacts with skeptical staff then it's multi-agent, and if we take into account the overall treatment then it's sequential.

Summary

- "Artificial intelligence" as an idea can be roughly classified along the dimensions thinking vs. acting and humanly vs. rationally. Acting rationally is what counts in applications.
- Recent successes with neural networks have changed the landscape of AI, moving many sub-areas more towards data-driven methods in the symbolic vs. subsymbolic divide. A grand challenge remains to find the best ways of combining the two kinds of methods.
- An agent is something that perceives and acts. A rational agent always takes the action that maximizes its expected performance, subject to the percept sequence and its environment knowledge.
- Some environments are more demanding than others . . .

  . . . your own, and that of James Bond, are the most difficult.

Reading

- *Chapter 1: Introduction* [Russell and Norvig (2010)].

  Content: A detailed account of the history of AI up to 2010.

- *Chapter 2: Intelligent Agents* [Russell and Norvig (2010)].

  Content: A more detailed description of concepts relating to agents, rationality, environments.

Elements of DSAI: Agenda

Week	#lectures	Topic	Lecturer(s)
1-2	3	Intro. DSAI	Dittrich, Hoffmann
3	2	Intro. Python	Dittrich
4-5	3	Intro. Machine Learning	Schiele
5-6	3	AlphaGo	Hoffmann
7-9	5	Autonomous Driving	Schiele
9-12	5	Dialogue Systems	Demberg
13-14	5	Big Data	Dittrich
15	1	Wrap Up	Dittrich, Hoffmann

Quiz: Rationality

Quiz

Who is rational?

(A): James Bond, crossing the street without looking.

(B): Your dog, crossing the street without looking.

(C): Vacuum cleaner, deciding to clean under your bed.

(D): Thermostat, deciding to cool down your fridge.

(A) Depends which role personal safety plays in his performance measure.

(B) Depends on whether or not dogs are able to check the traffic. If they can't, then this could be optimal (e.g. to meet fellow dogs or grab a sausage).

(C) Yes. (In case such percepts are available to the vacuum cleaner: only if it's dirty under your bed and you're not currently sleeping in it.)

(D) Not clear whether a thermostat is "taking actions" – is an "agent" – in a meaningful way. It's just a physical reaction (like a solar panel that produces electricity if the sun shines).

Quiz: Environments

Quiz: James Bond's environment is?

→ (A) Fully Observable: Definitely not! Else Bond would always know immediately what the bad guys are up to.

→ (B) Episodic: Definitely not. Every one of Bond's "actions" would be "rewarded" separately and independently. The "film plot" would consist of saving/not-saving the world every 2 minutes.

→ (C) Static: Definitely not. Just imagine Bond standing there, thinking, while the bad guys release Godzilla in Dillingen.

→ (D) Single-Agent: Definitely not. A Bond film without bad guys would be boring.

Quiz: Your own environment is?

→ (A) Fully Observable: No. E.g., you don't know what the exam questions will be.

→ (B) Episodic: No. E.g., it takes more than one action to complete your studies.

→ (C) Static: No. E.g., if you take a year to decide how to prepare for the exam, it'll be over by the time you're done.

→ (D) Single-Agent: No. Apart from your family etc., for example at some point you will compete for the same job with somebody else.

References I

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 1995.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (Third Edition)*. 2010.