Universität Münster

What's Intelligence Got To Do With It?

An Introduction into Artificial Intelligence Research

Tanya Braun

Data Science Group
Computer Science Department

Universität Münster

Acknowledgements

Slides are an amalgamation of books, articles, slides, and discussions with colleagues, most notably, by and with Ute Schmid, Subbarao Kambhampati, Stuart Russell, Malte Schilling, Ralf Möller, and Marcel Gehrke

Thank you!

living.knowledge

Data Science Group
Computer Science Department

Universität Münster

Broad Goals of the Talk Today

- We will talk about
  - History of AI: Knowledge-driven vs. data-driven AI
  - Approaches to AI: Thinking or acting humanly or rationally
  - Future of AI: human-centric and hybrid AI?
  - ✓ Get an idea of *where it's been, what it's doing & where it's going – maybe*

- This talk cannot provide
  - Complete overview of all the methods that fall under AI methods
  - Tutorial on how to use machine learning techniques for medicine
  - In-depth explanation of ChatGPT and Large Language Models (LLMs)

Universität Münster

Where it’s been

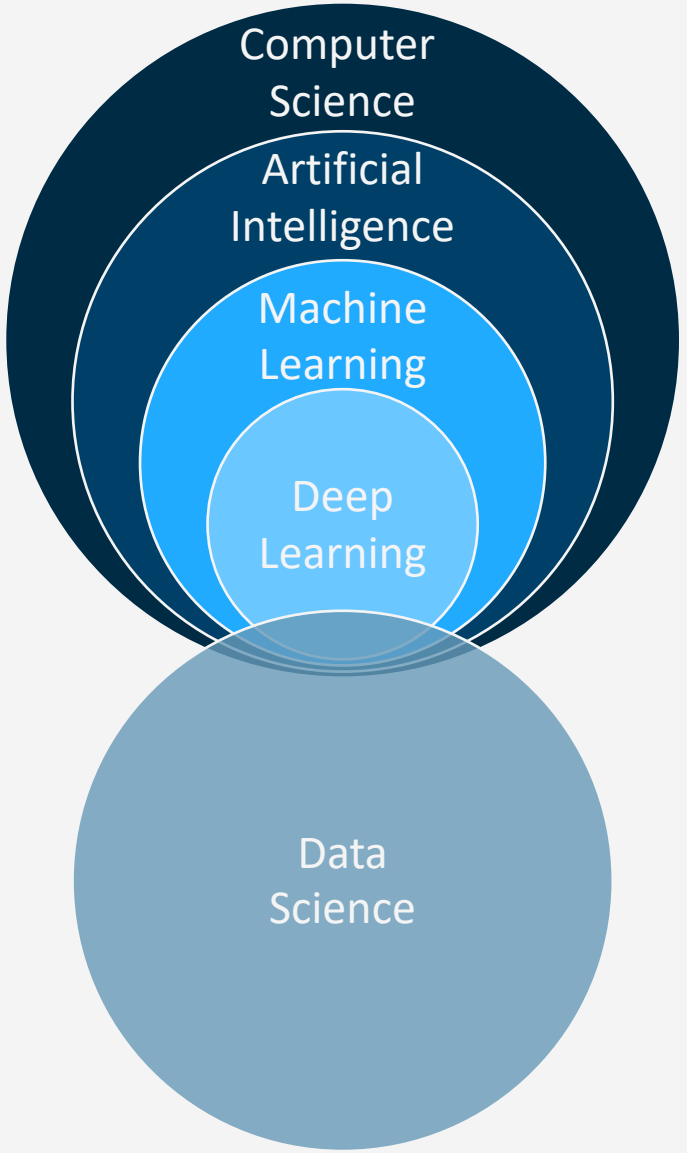
Artificial Intelligence Research

living.knowledge

Data Science Grouo
Computer Science Department

Artificial Intelligence (AI)

- Computer science field
  - Inception: 1956 (John McCarthy, Stanford)
  - Defined by McCarthy as "*the science and engineering of making intelligent machines*"
- Most computer programs do not rely on AI
- Using AI methods means giving up on completeness and correctness
  - Reasonable to use AI methods if
    - Problem so complex that optimal solution cannot be efficiently computed → heuristic methods, approximations
    - Problem cannot be (completely) specified → replace explicit algorithms with models / programs learned from data (black box)



A Bit of History: 1st Wave of AI

- Focus: Explicit knowledge representation
  - Also called intelligent design
    - Figure out what you want, encode knowledge explicitly in some representation, tell computer how to manipulate representation to get what you want
    - Started out logic-based
  - Constitutes powerful inference methods, provable properties, comprehensibility
- Problem: Polanyi's Paradox
  - "We know more than we can tell"
  - Large part of knowledge not verbalisable → only implicitly available
  - Focus on explicit knowledge tasks instead of tacit knowledge tasks
  - Brittle models
    - World too complex

Knowledge-driven or symbolic AI

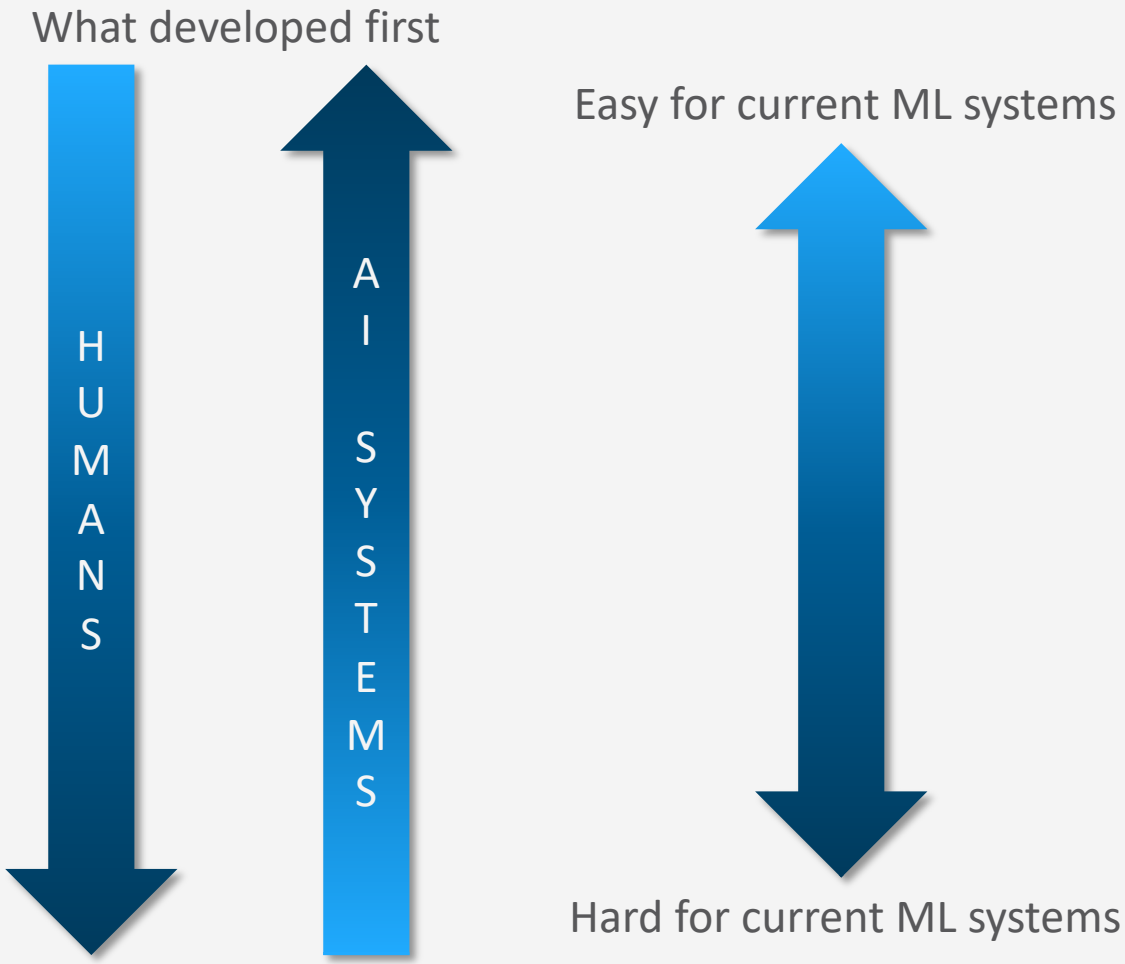
A Bit of History: 2nd Wave of AI

- Focus: Data-intensive machine learning
  - Use available data to learn a model representing tacit knowledge
    - Show the computer lots of examples of inputs together with the desired outputs. Let the computer learn how to map inputs to outputs using a general purpose, learning procedure
    - Took off around 2012 probably
  - Impressive results, especially in image-based classification
- But 1: Huge effort to get a large amount of high-quality data
  - Also applies here: *Garbage in – garbage out*
  - Especially a problem in highly specialised areas such as medical computer science
    - E.g., we are currently looking at a data set of 150 data points with 350,000 features each
- But 2: Limited explainability / comprehensibility of very complex models

Data-driven or neural AI

The Many Intelligences

- Perceptual & manipulation intelligence
  - Image recognition; hand-eye coordination
  - Largely tacit / subconscious
- Emotional intelligence
  - Showing & recognising emotional responses
- Social & communicative intelligence
  - Language
  - Requires a “theory of mind”
- Cognitive / reasoning intelligence
  - Hopefully, what we get tested for in uni
  - More declarative / consciously accessible



Subbarao Kambhampati: “Polanyi vs. Planning (Planning around AI’s New Romance with Tacit Knowledge)”, invited talk, DC, ICAPS 2020.

Why Did AI Develop in “Reverse”?

- It is easier to program computers on aspects of intelligence for which we have conscious theories (Polanyi’s Paradox)
  - Ergo the progress in reasoning / cognitive intelligence during the 1st wave of AI
- We are not particularly conscious of perceptual (and manipulative) intelligence
  - We had to depend on making machines learn the way we had to
    - Learn from data / demonstrations

Subbarao Kambhampati: “Polanyi vs. Planning (Planning around AI’s New Romance with Tacit Knowledge)”, invited talk, DC, ICAPS 2020.

Inference vs. Learning Focus

The Interpretability Issue:

If the representations are learned, how do we ensure that they are understandable to the humans?

Inference

- Start by assuming models available
  - State / action representation etc.
- Focus on inference in context of model
  - Promise of eventually learning / updating of models
    - Postpones learning; reasonable for explicit knowledge domains with good models (Chess, Sudoku, mission planning…)
- AI development followed this direction for much of its history

Learning

- Assume that the agent does not have any a priori models
- Focus on learning (even primitive) models
  - Typically, reflex agents
  - Promise of eventually getting to inference
    - Postpones inference; reasonable for tacit knowledge domains with no good models but a lot of examples / experience generators (vision, NLP, etc. …)
- Significant recent progress

Subbarao Kambhampati: “Polanyi vs. Planning (Planning around AI’s New Romance with Tacit Knowledge)”, invited talk, DC, ICAPS 2020.

Universität Münster

What it’s doing

Artificial Intelligence Research

living.knowledge

Data Science Grouo
Computer Science Department

A Bouquet of AI Methods

- Problem-solving
  - Search algorithms, heuristics, game theory, constraint satisfaction problems, …
- Logic
  - Propositional logic, description logic, ontologies, knowledge graphs; inference
- Uncertainty
  - Probabilistic modelling and inference (over time), utility and decision theory, multi-agent systems
- Machine learning
  - Learning from examples; neural networks, deep learning, reinforcement learning, …
- Perceiving and acting
  - Natural language processing, computer vision, robotics

Approaches to Artificial Intelligence (AI)

- All approaches researched
  - Supported and hindered each other
- Rationality
  - System is rational if it does the "right thing," given what it knows

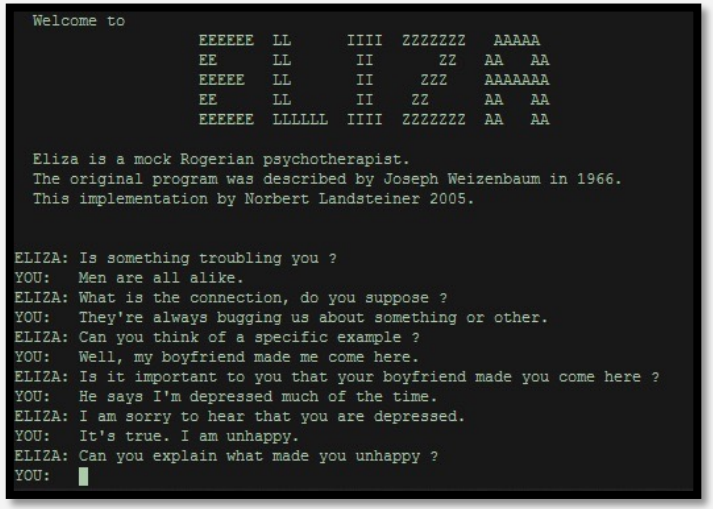
Success measure

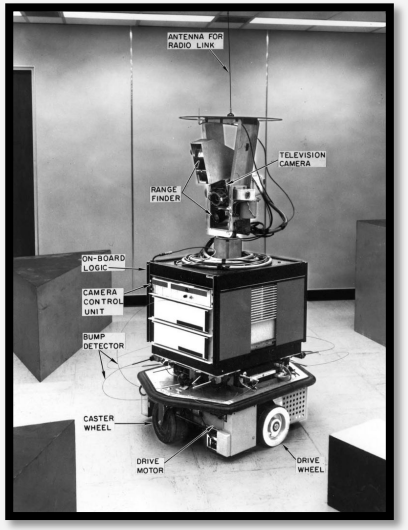
Fidelity of human performance	Ideal performance measure rationality	
Thinking Humanly	**Thinking Rationally**	Thought processes, reasoning
"The exciting new effort to make computers think . . . machines with minds, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning…" (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)	
Acting Humanly	**Acting Rationally**	Behaviour
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI … is concerned with intelligent behaviour in artefacts." (Nilsson, 1998)	

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Acting Humanly

- Turing Test (Turing, 1950)
  - Computer passes test, if a human, who asks written questions, cannot tell if the the written answers come from a human or not
    - Example: *Eliza*, program for superficially simulating a psychiatrist
    - See also Ch. 26, “Artificial Intelligence – A Modern Approach” by Russel & Norvig, including a discussion whether a computer would really be intelligent if it passed
      - Regarding Eliza: human’s example closure tendencies are more pronounced for emotional/social intelligence aspects
      - Cf. robot *Shakey*: No on who saw Shakey the first time thought it could shoot hoops, yet the first people interacting with Eliza assumed it was a real doctor
  - *Total* Turing Test: includes a video signal to test perceptual abilities, opportunity to pass physical objects





https://en.wikipedia.org/wiki/ELIZA
https://thenewstack.io/remembering-shakey-first-intelligent-robot/

Acting Humanly

- Subproblems to solve as part of the Turing Test
  - *Natural Language Processing*
    - Communication
  - *Knowledge representation*
    - Store knowledge and observations
  - *Automated reasoning*
    - Answer questions, draw new conclusions
  - *Machine learning*
    - Adapt to new circumstances, detect and extrapolate patterns
  - Total Turing Test
    - *Computer vision*: perceive objects
    - *Robotics*: manipulate objects, move about

The Turing Test covers a majority of disciplines that make up AI nowadays.

- But:
  little research effort devoted to pass test
- Instead:
  Study underlying principles of intelligence

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Approaches to Artificial Intelligence (AI)

- All approaches followed
  - Supported and hindered each other
- Rationality
  - System is rational if it does the "right thing," given what it knows

Success measure

	Fidelity of human performance	Ideal performance measure rationality
Thought processes, reasoning	**Thinking Humanly** "The exciting new effort to make computers think . . . machines with minds, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning…" (Bellman, 1978)	**Thinking Rationally** "The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Behaviour	**Acting Humanly** "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	**Acting Rationally** "Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI … is concerned with intelligent behaviour in artefacts." (Nilsson, 1998)

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Thinking Humanly

- A “program thinks like a human“
  - Requires a way to determine how humans think → workings of the human mind
  - Given theory of the mind, express theory as computer program
    - If program’s input-output behaviour matches corresponding human behaviour, evidence that some of program’s mechanisms could also be operating in humans
- Approach complementary to AI: *Cognitive Science*
  - Interdisciplinary:
    - Computer models from AI
    - Experimental techniques from psychology
  - Goal:
    Construct precise and testable theories of human mind

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Approaches to Artificial Intelligence (AI)

- All approaches followed
  - Supported and hindered each other
- Rationality
  - System is rational if it does the "right thing," given what it knows

Success measure

	Fidelity of human performance	Ideal performance measure rationality
	Thinking Humanly	**Thinking Rationally**
	"The exciting new effort to make computers think . . . machines with minds, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning…" (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
	Acting Humanly	**Acting Rationally**
	"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI … is concerned with intelligent behaviour in artefacts." (Nilsson, 1998)

Thought processes, reasoning

Behaviour

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Thinking Rationally

- Codify thinking → rules
  - Irrefutable reasoning processes
  - Argument structures that always yield correct conclusions when given correct premises
- Field of *Logic*
  - Precise notation for statements about objects in a world and relations among them
  - Programs that could, in principle, solve *any* solvable problem described in logical notation
  - Obstacles:
    - Informal knowledge
      - Unstructured data
      - Uncertainty
    - Solving any solvable problem in practice
      - Limited computational resources

Obstacles apply to *any* attempt to build computational reasoning systems
- Formulated first in logic

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Approaches to Artificial Intelligence (AI)

- All approaches followed
  - Supported and hindered each other
- Rationality
  - System is rational if it does the "right thing," given what it knows

Success measure

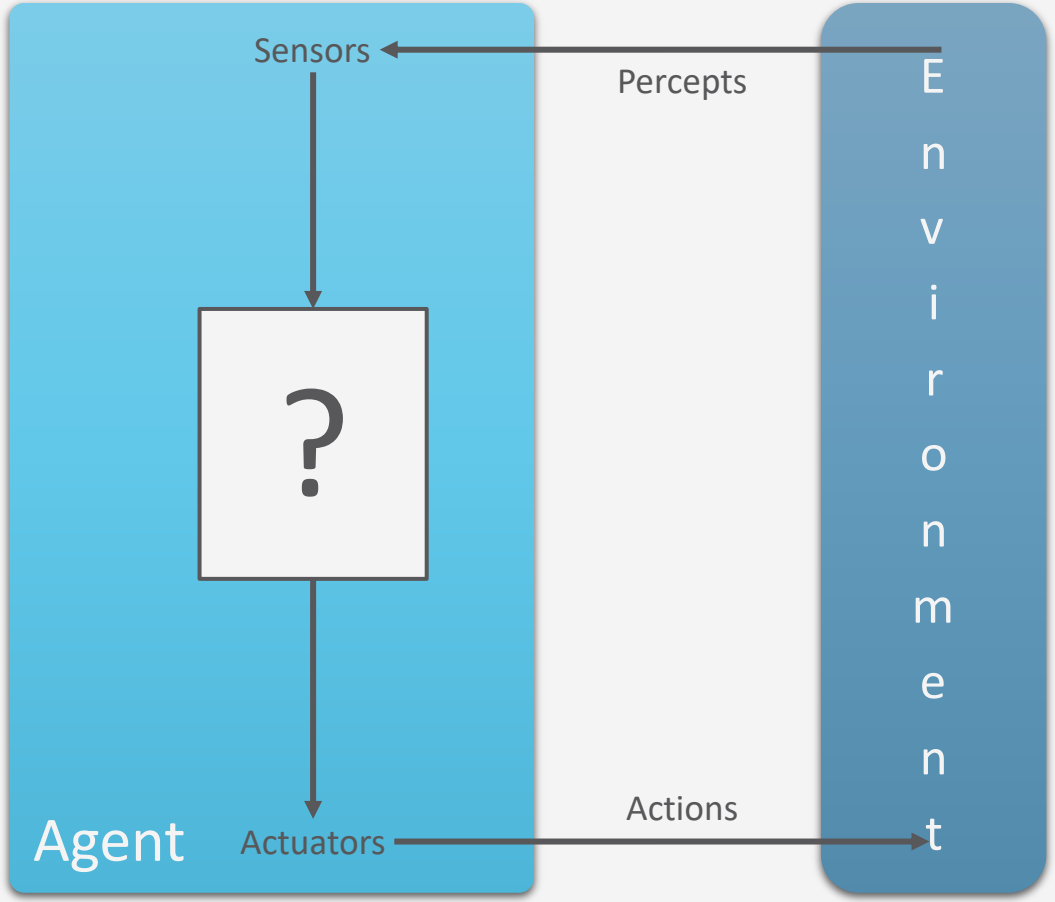
	Fidelity of human performance	Ideal performance measure rationality
Thought processes, reasoning	**Thinking Humanly** "The exciting new effort to make computers think . . . machines with minds, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning…" (Bellman, 1978)	**Thinking Rationally** "The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Behaviour	**Acting Humanly** "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	**Acting Rationally** "Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI … is concerned with intelligent behaviour in artefacts." (Nilsson, 1998)

S. Russel and P. Norvig: Artificial Intelligence – A Modern Approach, 2020.

Acting Rationally

Advantage: Standard of rationality mathematically well defined
- Better suited to generate agent designs that provably achieve rationality

- Rational agent approach
- Agent = something that acts
  - Operates autonomously
  - Perceives environment
  - Persists over a prolonged time period
  - Adapts to change
  - Creates and pursues goals
- Rational agent
  - Acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome
  - May include thinking rationally or acting humanly, but *more general*

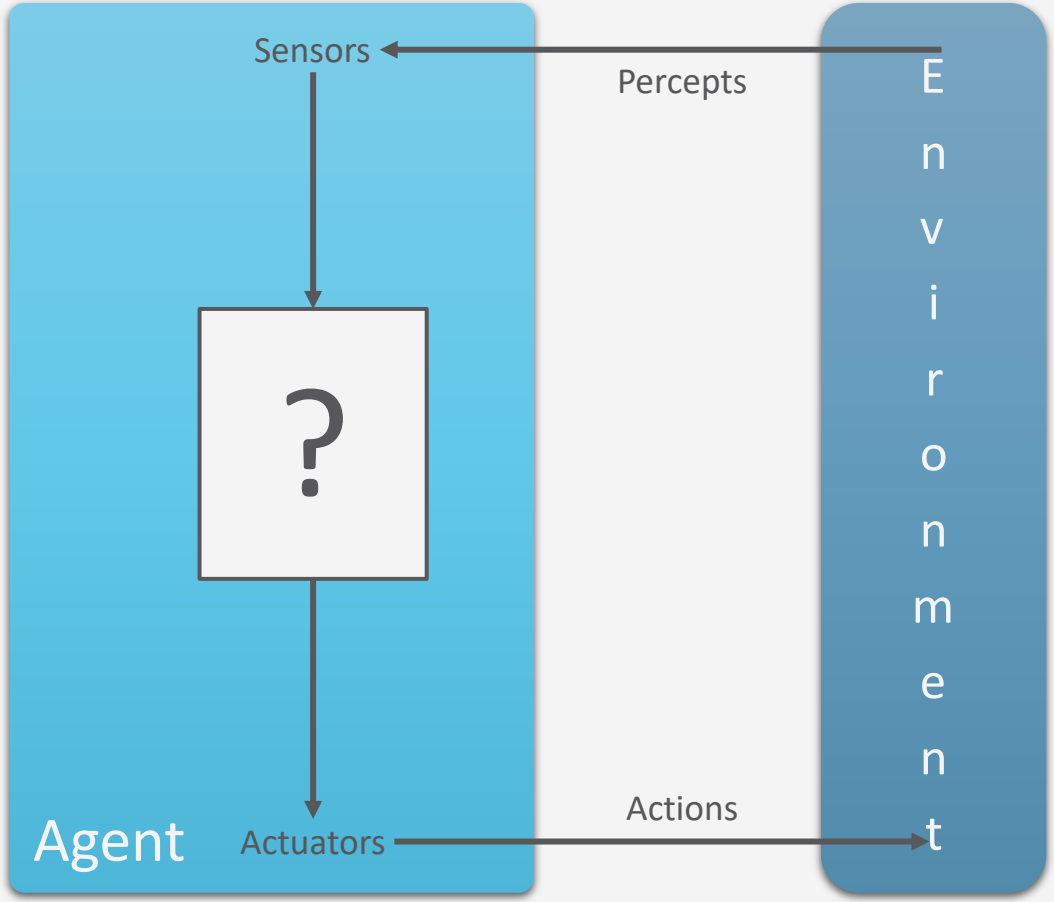


Rationality

Rationality is not omniscience!

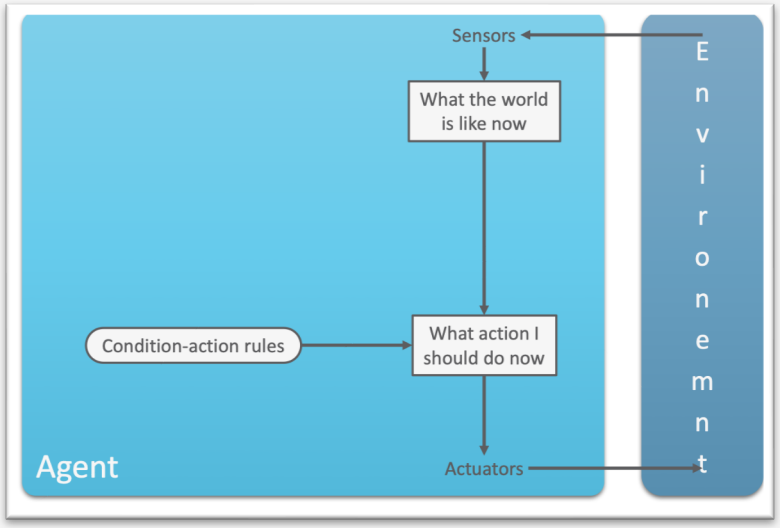
- Depends on four things:
  - *Performance measur*e, defines criterion of success
  - Agent's prior *knowledge* of environment
  - *Actions* that agent can perform
  - Agent's *percept sequence* to date
- Rational agent:
  - For each possible percept sequence, a rational agent should select an *action*
  - expected to maximize its *performance measure*,
  - given evidence provided by *percept sequence* and
  - whatever built-in *knowledge* the agent has.

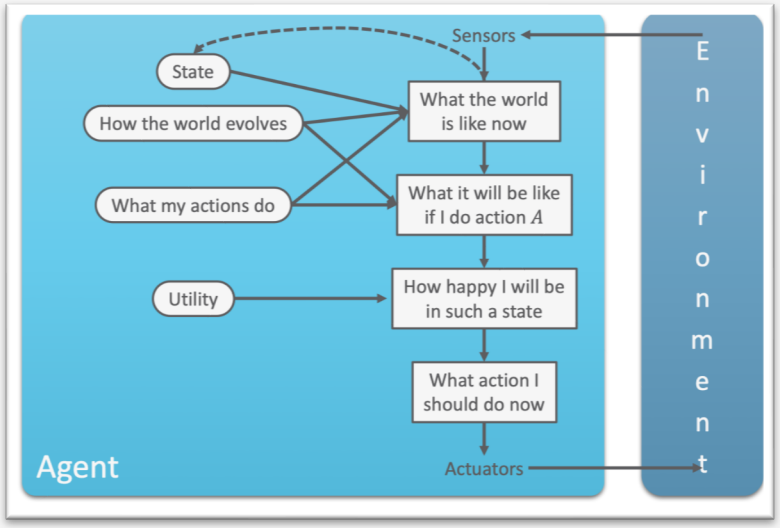
→ Rational = intelligent

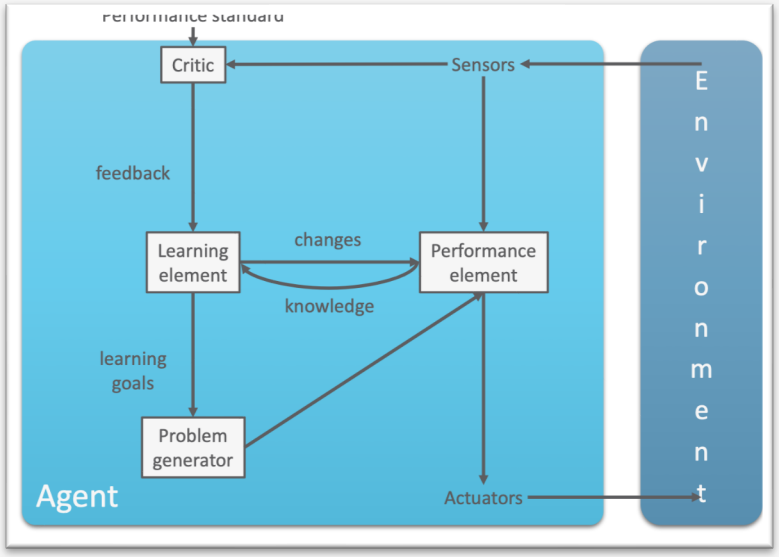


Agent Structure

- Implementation can become arbitrarily complex
  - Left: reflex agent, middle: model-based utility-based agent, right: learning agent







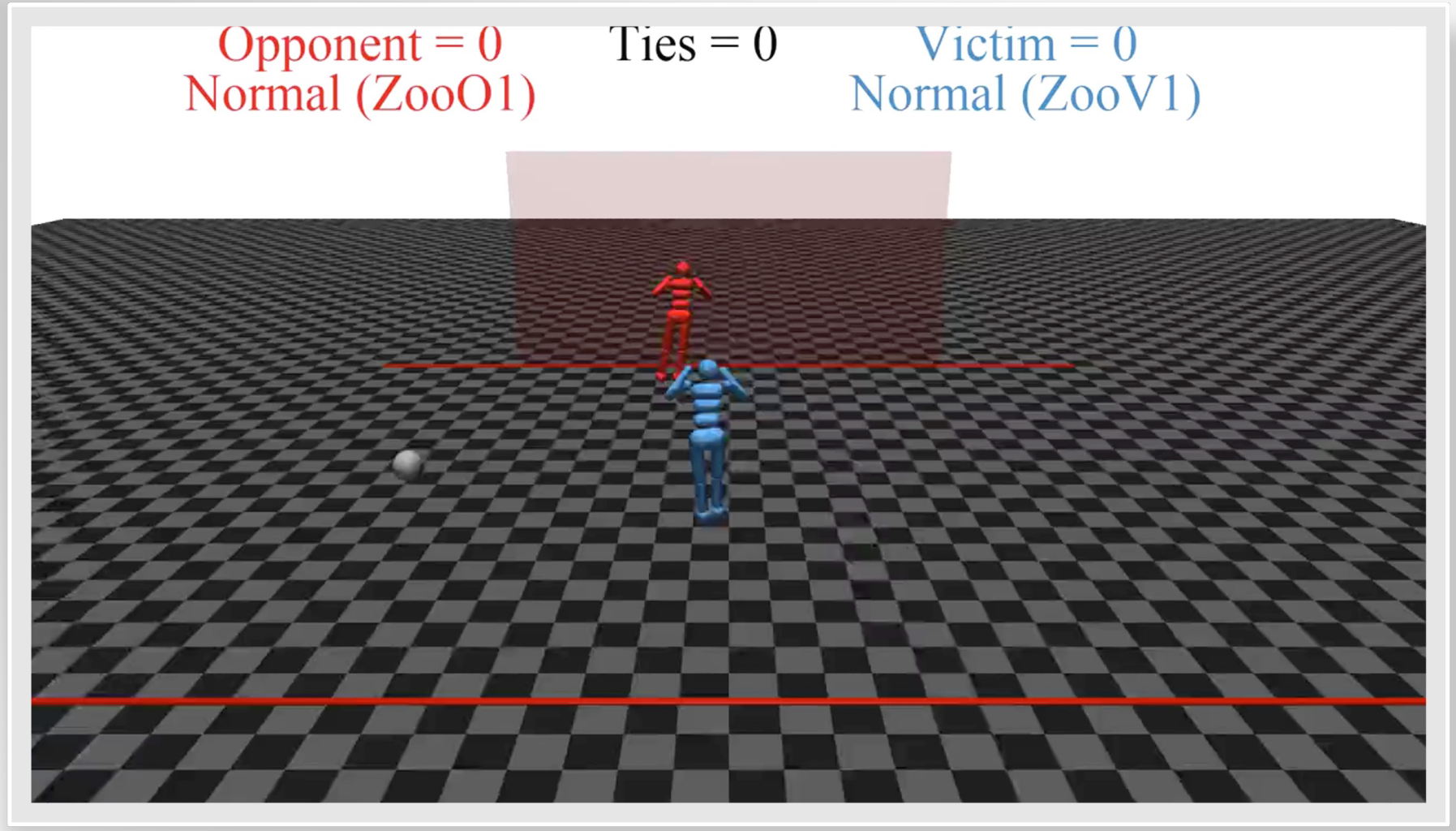
Universität Münster

Where it’s going – maybe

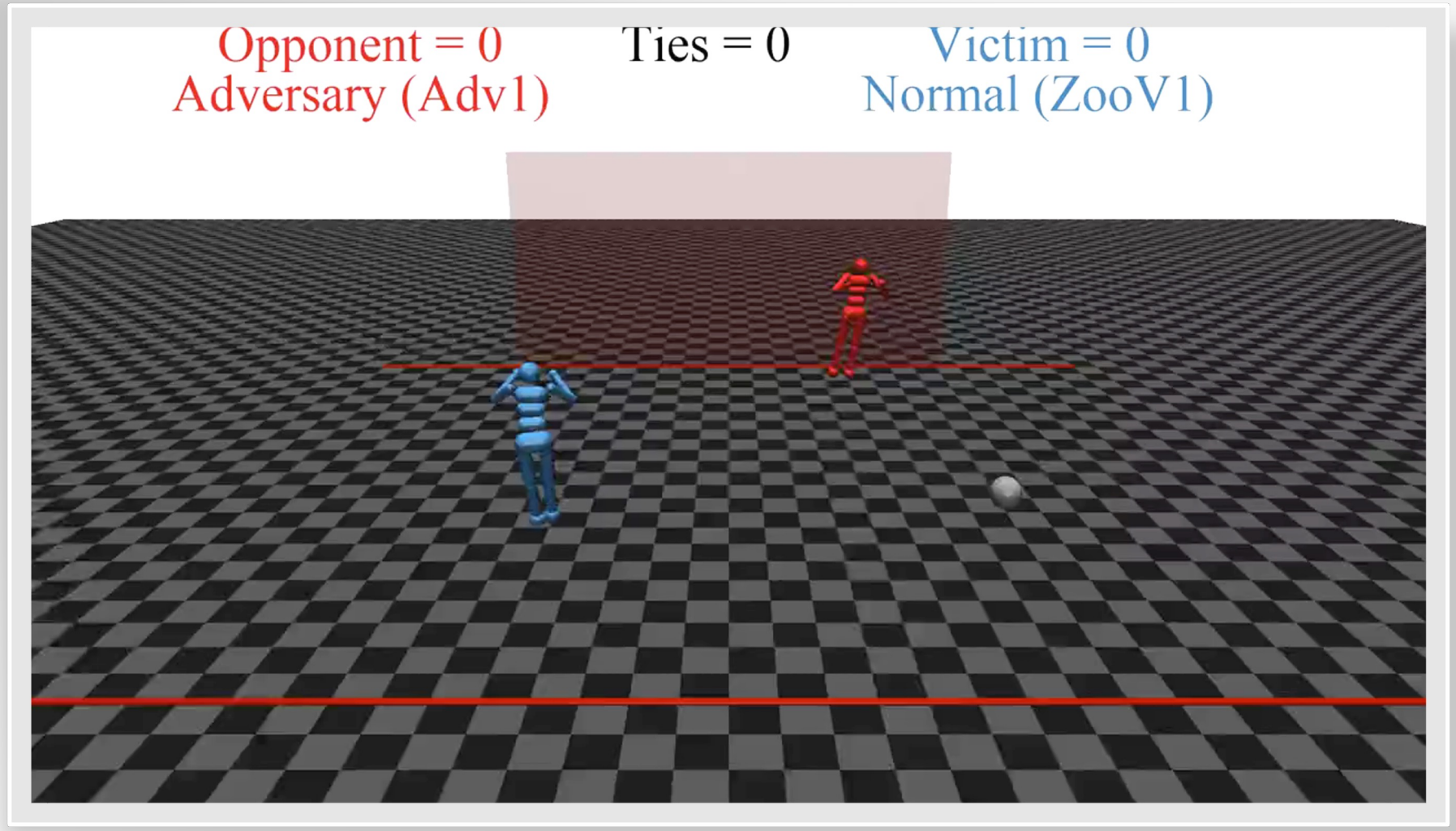
Artificial Intelligence Research

living.knowledge

Data Science Grouo
Computer Science Department



Quelle: https://people.eecs.berkeley.edu/~russell/talks/2020/russell-aaai20-hntdtwwai-4x3.pptx



Quelle: https://people.eecs.berkeley.edu/~russell/talks/2020/russell-aaai20-hntdtwwai-4x3.pptx

A Bit of the Future: 3rd Wave of AI?

- Focus: human-centred AI, hybrid approaches
  - Human-centred: Human-aware AI, explainable AI (XAI), interactive machine learning
    - Different dimensions
      - Human as a source for training
      - Human for which outputs should be comprehensible
      - Human and system working as a team
  - Hybrid: Combine data-driven and knowledge-driven approaches
    - Also known as *neuro-symbolic*
    - Use knowledge during learning to combat the problem of requiring a huge amount of data

Polanyi's Revenge

Kambhampati, Subbarao. "Polanyi's Revenge and AI's New Romance with Tacit Knowledge". In *Communications of the ACM*, 2021.

Hybrid / human-centred AI

Universität
Münster

The Finish Line

Artificial Intelligence Research

living.knowledge

Data Science Grouo
Computer Science Department

What's intelligence got to do with it?

Where it's been

- Knowledge-driven AI: model-based inference, provable properties, comprehsenibility → brittle!
- Data-driven AI: learn a model from huge amounts of input-output pairs → interpretability issue!

What it's doing

- AI methods: search-based problem solving, logic-based inference and knowledge representation, probabilistic modelling and reasoning under uncertainty, machine learning, perception and action

Where it's going – maybe

- Hybrid AI: combine knowledge- and data-driven AI methods
- Human-centric AI: Do not forget the human in all of this!
  - And all the things that come with it:
    Ethics, robustness, safety, transparency, trustworthiness, …

Thank you!