Einführung in die Künstliche Intelligenz

- Dozenten
  - Prof. Michael Schroeder
  - Prof. Carsten Rother

- Labs
  - Dr. Daniel Eisinger
  - Dr. Dimitrij Schlesinger

- URL: teaching on
  www.biotec.tu-dresden.de/schroeder

- thanks to Prof. Fürnkranz, Darmstadt for slide templates
Text Book

  - (original: AI – a modern approach)

- Course slides
  - in English (lecture is in German)
  - will be available from Home-page
What is Artificial Intelligence

- Different definitions due to different criteria
- Two dimensions:
  - Thought processes/reasoning vs. behavior/action
  - Success according to human standards vs. success according to an ideal concept of intelligence: rationality.

<table>
<thead>
<tr>
<th>Systems that <strong>think</strong> like humans</th>
<th>Systems that <strong>think</strong> rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems that <strong>act</strong> like humans</td>
<td>Systems that <strong>act</strong> rationally</td>
</tr>
</tbody>
</table>
## Definitions of Artificial Intelligence

<table>
<thead>
<tr>
<th>Systems that think like humans</th>
<th>Systems that think rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The exciting new effort to make computers think ... machines with minds, in the full and literal sense.” (Haugeland, 1985)</td>
<td>“The study of mental faculties through the use of computational models.” (Charniak and McDermott, 1985)</td>
</tr>
<tr>
<td>“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ...” (Bellman, 1978)</td>
<td>“The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Systems that act like humans</th>
<th>Systems that act rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)</td>
<td>“Computational Intelligence is the study of the design of intelligent agents.” (Poole et al., 1998)</td>
</tr>
<tr>
<td>“The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)</td>
<td>“AI ... is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)</td>
</tr>
</tbody>
</table>

**Figure 1.1** Some definitions of artificial intelligence, organized into four categories.
Systems that think like humans

- How do humans think?
- Requires scientific theories of internal brain activities (cognitive model):
  - Level of abstraction? (knowledge or circuitry?)
  - Validation?
    - Predicting and testing human behavior
    - Identification from neurological data
- Cognitive Science
- brings together computational models from AI
- and experimental techniques from psychology
- to construct precise and testable theories of the mind
- Cognitive Science is now distinct from AI
- Cognitive Neuroscience
- How does the brain work at the neuronal level?
Systems that think rationally

- Capturing the laws of thought
  - Aristotle: What are ‘correct’ argument and thought processes?
    - Correctness depends on irrefutability of reasoning processes.
    - Syllogisms:
      - first patterns of correct formal reasoning
      - “Socrates is a man, all man are mortal → Socrates is mortal”
  - This study initiated the field of logic.
    - The logicist tradition in AI hopes to create intelligent systems using logic programming.

- Problems:
  - It is hard to formalize knowledge exactly
    → Feigenbaum Bottleneck in expert systems
  - Practical constraints
    - it is easy to write a logically optimal chess player...
  - Not all intelligence is mediated by logic behavior
Systems that act like humans

- When does a system behave intelligently?
  - Turing (1950) Computing Machinery and Intelligence
    Operational test of intelligence: imitation game

- Test still relevant now, yet might be the wrong question.
- Requires the collaboration of major components of AI:
  knowledge, reasoning, language understanding, learning, …

- Problem with Turing test:
  - not reproducible, constructive or amenable to mathematical analysis.
Systems that act rationally

- Rational behavior: “doing the right thing”
  - The “Right thing” is that what is expected to maximize goal achievement given the available information.
- Can include thinking, yet in service of rational action
  - Action without thinking: e.g. reflexes.

- Two advantages over previous approaches
  - More general than law of thoughts approach
    - in many situations, a provably correct action does not exist
  - More amenable to scientific development
    - rationality can be defined and optimized

- On the other hand
  - perfect rationality is only feasible in ideal environments.
  - rationality is often not a very good model of reality.
    - humans are, e.g., very bad in estimating probabilities...
Foundations of AI

Different fields have contributed to AI in the form of ideas, viewpoints and techniques.

- **Philosophy**: Logic, reasoning, mind as a physical system, foundations of learning, language and rationality.
- **Mathematics**: Formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability.
- **Psychology**: Adaptation, phenomena of perception and motor control.
- **Economics**: Formal theory of rational decisions, game theory.
- **Linguistics**: Knowledge representation, grammar.
- **Neuroscience**: Physical substrate for mental activities.
- **Control theory**: Homeostatic systems, stability, optimal agent design.
Subdisciplines of AI

- Natural Language Processing
- Knowledge Representation
- Automated Reasoning
- Planning
- Machine Learning
- Computer Vision
- Robotics
- ...

Einführung in die Künstliche Intelligenz
Leibniz (1646-1716): Calculus Raciocinator

Idea: Reasoning = prime factor computation
   concepts=numbers
   basic=prime
   complex=non-prime composed of basics’ primes

Example

animal=2, rational=3, therefore human=2x3=6
Assuming monkey=10 he concludes: monkey \neq human
because neither is 10/6 nor 6/10 int.

To prove the usefulness of his calculus, he assumes “those marvelous characteristic numbers” as given
A Brief History of AI

- 18th century:
  - von Kempelen's chess-playing Turk
  - amazing piece of mechanical engineering
  - played and won chess games all over the world (e.g., against Napoleon)
  - unfortunately a hoax...

- 19th century:
  - Charles Babbage’s Analytical Engine
  - DeMorgan and Boole
George Boole (1815-1864) Proposition I.

All the operations of Language, as an instrument of reasoning may be conducted by a system of signs composed of the following elements, viz:

1-st. Literal symbols, as x, y, etc., representing things as subject of our conceptions.

2-nd. Signs of operations, as +, -, x, standing for those operations of the mind by which the conceptions of things are combined or resolved so as to form new conceptions involving the same elements.

3-rd. The sign of identity, =.

And these symbols of Logic are in their use subject to definite laws, partly agreeing with and partly differing from the laws of the corresponding symbols in the science of Algebra.
Idea behind approach

Notion of class central: To get “good men” mind selects objects from class “men” with additional property “good”

Order does not affect selection (commutative and distributive law)

With arithmetics stating 0\(y\)=0 and 1\(y\)=1, he defines 0 is nothing and 1 the universe

First he uses fixed universe, but then uses De Morgan’s idea and thus implicitly requires 0=\(\neq\)=1
Boole’s Algebra

Commutative Law: \( x + y = y + x \quad xy = yx \)

Associative Law: \( (x + y) + z = x + (y + z) \)
\( (xy)z = x(yz) \)

Neutral Element: \( x + 0 = x \quad x1 = x \)

(Additive) Inverse: \( x + (-x) = 0 \)

Distributive Law: \( x(y + z) = xy + xz \)

\( x^2 = x \)

\( 0 \leq 1 \)

\( x + x = 0 \) implies \( x = 0 \)
Example

Definition of wealth due to the economist N.W. Senoir:
```
Wealth consists of things transferable, limited in supply, and either productive of pleasure or preventative of pain.
```

\[
w = st(p(1-r)+r(1-p))
\]

w representing wealth,
t transferable things,
s limited in supply,
p productive of pleasure and
r preventative of pain
A Brief History of AI

1943:
- McCulloch and Pitts model artificial neurons
- 1951: Marvin Minsky and Dann Edmonds constructed the first neural network computer

1950
- Claude Shannon: algorithm for playing Chess
- foresees strategies that are still used today
  - Shannon Type-A Strategy:
    - brute-force minimax search until a fixed horizon
  - Shannon Type-B Strategy
    - pruning uninteresting lines (as humans do)
    - preferred by Shannon (and contemporaries)

1951
- Turing's chess algorithm
  - computation with paper and pencil
  - first recorded man-machine (chess) game
The Dartmouth Conference

1956:

- John McCarthy invites 10 scientists with various backgrounds to a 2-week workshop at Dartmouth College.
  - Bringing together top minds on automata theory, neural nets and the study of intelligence.
  - For the next 20 years the field was dominated by these participants.
    - John McCarthy, Herbert Simon, Allan Newell, Marvin Minsky, Arthur Samuel, etc.
  - Allen Newell and Herbert Simon (CMU): The Logic Theorist
    - First nonnumerical thinking program used for theorem proving.
    - Proved various theorems of Whitehead's *Principia Mathematica*.
    - A joint publication by AN, HS, and LT was rejected...

→ Term “Artificial Intelligence” is coined
Great Expectations
(1952-1969)

- Newell and Simon (CMU): the General Problem Solver.
  - Imitation of human problem-solving
  - successfully solved simple puzzles
  - physical symbol system hypothesis
- Arthur Samuel (IBM, 1952-)
  - investigated game playing (checkers) with great success at IBM
    - program beat a regional master
    - pioneered many ideas in game playing and machine learning
      - including alpha-beta search, reinforcement learning, etc.
- John McCarthy (MIT, 1958-)
  - Inventor of Lisp (second-oldest high-level language)
  - Logic-oriented Advice Taker
    - separation between knowledge and reasoning
- Marvin Minsky (MIT)
  - various students working on micro-worlds (e.g., block's world)
A Dose of Reality (1966-1973)

- Progress was slower than (unrealistic) expectations
  - Simon and Newell's (1958) predictions

1. That within ten years, a digital computer will be the world’s chess champion, unless the rules bar it from competition.
2. That within ten years a digital computer will discover and prove an important new mathematical theorem.
3. That within ten years a digital computer will write music that will be accepted by critics as possessing considerable aesthetic value.
4. That within ten years most theories in psychology will take the form of computer programs, or of qualitative statements about the characteristics of computer programs.
   ...The simplest way I can summarize is to say that there are now in the world machines that can think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until — in a visible future — the range of problems they can handle will be coextensive with the range to which the human mind has been applied.
A Dose of Reality 
(1966-1973)

- Progress was slower than (unrealistic) expectations
  - Simon and Newell's (1958) predictions
  - came more or less true after 40 (instead of 10) years
    - but in very different ways than they had imagined

- Difficulties
  - Difficulty of knowledge representation
    - e.g., attempt for Machine Translation of Russian Scientific papers
      - the spirit is willing but the flesh is weak
        → the vodka is good but the meat is rotten

- In 1966 no machine translations were used
- nowadays they are routinely used
  - but most still still give awful results
From: bemerk@tu-darmstadt.de<bemerken@tu-darmstadt.de>
Date: 2012/9/21
Subject: Mail: Reagieren Sofort [MH-Hannover]
To:

KLICKEN ANTWORT VOR FÜLLEN INFORMATIONEN

für Ihre Informationen,


Bitte tun nicht ignorieren.

MH-Hannover
A Dose of Reality (1966-1973)

- Progress was slower than (unrealistic) expectations
  - some of Simon and Newell's (1958) predictions came more or less true after 40 (instead of 10) years
    - but in very different ways than they had imagined
- Difficulties
  - Difficulty of knowledge representation
  - Lack of scalability
    - under-estimation of the combinatorial explosion in search
    - things that work well in micro-worlds do not work in real world
      - e.g., theorem proving only worked with very few facts
    - Lighthill report (1973) focused on this issue
      - stopped AI funding in UK in all but two universities
  - Fundamental limitations on techniques and representations
    - Minsky and Papert (1969) noted that perceptrons are only linear separators
      - killed research in neural networks for decades
Knowledge-Based Systems (1969-1979)

- DENDRAL project (Buchanan et al. 1969)
  - task:
    - infer molecular structure from formula of the molecule and the mass spectrum
  - First successful knowledge-intensive system
    - recognized importance of domain-specific knowledge

- Expert systems
  - MYCIN to diagnose blood infections (Feigenbaum et al.)
    - with ~450 rules, it performed better than junior doctors
    - knowledge had to be tediously acquired from experts
      (→ Knowledge Engineering Bottleneck)
    - introduction of uncertainty in reasoning

- Increase in knowledge representation research
  - Logic, frames, Schank's scripts, semantic nets, …
  - CYC project (started by Lenat 1984)
    - attempt to encode common-sense knowledge
The AI industry (1980-present)

- R1 (McDermott, 1982)
  - expert system for configuring computers at DEC
  - saved about $40 million a year in 1986
- Fifth generation project in Japan (1981)
  - 10-year plan with strong focus on Logic Programming
  - did not quite live up to its ambitious goals
- Similar programs in US and UK put an end to the AI winter
  - in UK, the Alvey report reinstated funding for “Intelligent Knowledge-Based Systems” (to avoid the name AI)
- Neural Networks revival
  - seminal work *Parallel Distributed Processing* by Rumelhart and McClelland (1986)
  - soon became popular in industrial applications
- AI industry grew from a few million dollars in 1980 to billions of dollars in 1988
Current Trends in AI

- exploit the strengths of the computer
  - fast repetitious computations
  - elementary statistics (counting and probabilities)
- and do not try to model human thought processes
  - brute-force methods perform much better in many areas
- aviation was only possible when planes stopped to flap their wings...
- focus on particular tasks and not on solving AI as a whole
  - “Intelligent prostheses”
    - tools that support us in tasks that would otherwise require human intelligence
  - “AI-complete problems”
    - informal notion for very hard problems that cannot be solved unless the system has human-like knowledge and reasoning
- strong scientific standards developed in the 1980s
  - solid experimentation and scientific verification of hypotheses
The Science of AI

- Conference Series
  - 1969 - : Biennial International Joint Conference on AI (IJCAI)
  - 1980 - : Annual National Conference on AI (AAAI)
  - 1982 - : Biennial European Conference on AI (ECAI)

- Magazines
  - AI Magazine (published by AAAI)
  - IEEE Intelligent systems

- Journals
  - Artificial Intelligence (Elsevier)
  - Journal of Artificial Intelligence Research
    - pioneered free on-line publication ([http://www.jair.org](http://www.jair.org))

- Since the 1980s various subfields emerged, joined forces with related fields
  - many journals and annual conferences in subareas
State of the Art

- Autonomous Planning and Scheduling
  - NASA's on-board autonomous planning program controlled the scheduling of operations for a spacecraft
  - During the 1991 Gulf War, US forces deployed an AI logistics planning and scheduling program
    - DARPA: this application paid back 30 years of investment in AI

- Game Playing
  - TD-Gammon learned an evaluation function for backgammon that led to changes in backgammon theory
  - Deep Blue defeated the reigning world chess champion Garry Kasparov in 1997
    - increased IBM stocks by several billion $

- Spam Filtering
  - machine learning techniques are commonly used to adapt to new types of spam
State of the Art

- **Robotics**
  - No hands across America
    - CMU's ALVINN drives autonomously 98% of the time from Pittsburgh to San Diego
    - 2% human intervention mostly on exit ramps etc.
  - DARPA Grand Challenge
    - 2005 off-road race for autonomous vehicles
    - 2007 Urban Challenge
  - robot assistants are routinely used in microsurgery
  - over 2,000,000 Roomba vacuum robots sold

- **Scientific Discovery**
  - machine learning system helped to discover new quasars
  - automated theorem prover proved Robbins conjecture unsolved for decades
  - Data Mining and Knowledge Discovery has developed into a new industry
State-of-the-Art

- Natural Language Understanding
  - translation systems are frequently used
  - Speech recognition
    - United Airlines has an entire booking system based on automated speech recognition and dialog management

- Puzzles
  - Proverb solves crossword puzzles better than most humans
    - although it does not “understand” the clues
    - combination of constraint satisfaction, statistics, web retrieval, ...

- IBM's Watson has won at Jeopardy quiz show
  - http://www.ibm.com/de/watson