2025-05-01

This document contains the administrative information and overview chapter of the course notes for the course "Artificial Intelligence 2" held at FAU Erlangen-Nürnberg in the Summer Semesters 2017 ff.

Other parts of the lecture notes can be found at `http://kwarc.info/teaching/AI/notes-*.pdf`.

Contents

21.1 Preliminaries

In this section, we want to get all the organizational matters out of the way, so that we can get course contents unencumbered. We will talk about the necessary administrative details, go into how students can get most out of the course, talk about where the various resources provided with the course can be found, and finally introduce the ALEA system, an experimental – using AI methods – learning support system for the AI-2 course.

What you should learn here...

- **What you should learn in AI-2:**
  - In the broadest sense: *A bunch of tools for your toolchest* (i.e. various (quasi-mathematical) models, first and foremost)
  - the underlying *principles* of these models (assumptions, limitations, the math behind them ...)
  - the ability to describe real-world problems in terms of these models, **where adequate** (...and knowing **when** they are adequate!), and
  - the ideas behind effective *algorithms* that solve these problems (and to understand them well enough to implement them)
- **Note:** You will likely never get payed to implement an algorithm that e.g. solves Bayesian networks. (They already exist)
  - *But* you might get payed to *recognize* that some given problem *can be* represented as a Bayesian network!
  - **Or:** you can recognize that it is *similar to* a Bayesian network, and reuse the underlying principles to develop new specialized tools.

In other words: Many things you learn here are *means to an end* (e.g. understanding the underlying *ideas* behind algorithms), not the end itself. But the best way to understand these means is to first treat them as an end in themselves.

Compare two employees

- **"We have the following problem and we need a solution: ..."**
- **Employee 1 – Deep Learning can do everything:** "I just need ≈1.5 million labeled examples of potentially sensitive data, a GPU cluster for training, and a few weeks to train, tweak and finetune the model.

  But *then* I can solve the problem... with a confidence of 95%, within 40 seconds of inference per input. Oh, as long as the input isn't longer than 15`unit`, or I will need to retrain on a bigger input layer..."
- **Employee 2 – AI-2 Alumna:** "...while you were talking, I quickly built a custom UI for an off-the-shelve `<problem>` solver that runs on a medium-sized potato and returns a *provably correct* result in a few milliseconds. For inputs longer than 1000`unit`, you might need a slightly bigger potato though..."

▷ **Moral of the story:** Know your *tools* well enough to select the right one for the job.

Obviously, that is not to say that machine learning is not a useful tool! (It is!)

If your job is to e.g. filter customer support requests, or to recognize cats in pictures, trying to write a prolog program from scratch is probably the wrong approach: Just use a language model / image model and finetune it on a classification head.

But it is also not the only tool, and it is not always the right tool for the job – despite what some people might tell you. And even in scenarios where machine learning *can* yield decent results, it is not always the *best* tool. (Some people care about efficiency, explainability, etc ;))

In an ideal world . . . We would spend weeks on each topic, give you lots of interesting problems to solve, give you individual feedback and tutoring.

As an exam, you would have to solve a few real-world problems by choosing the right tools, model the problem accordingly, customize the algorithms to the specifics, implement them.

⇝ You would each write a 10 page essay in 4 hours, we would spend the next 6 months grading them, and then 95% of you would probably fail: Really understanding this stuff takes time and *lots of practice!*

Instead: we will teach you all the important stuff, give you practice problems to do on your own, and then test you on the basics in a manner that is actually gradable in a reasonable time frame, and doable

Hopefully, in five years, when you encounter a problem, you will remember enough of the broad strokes to recognize the "kind of problem" you have, and are able to look up the rest easily.

21.1.1 Administrative Ground Rules

We will now go through the ground rules for the course. This is a kind of a social contract between the instructor and the students. Both have to keep their side of the deal to make learning as efficient and painless as possible.

Prerequisites

- ▷ **Remember:** **AI-1** dealt with situations with "complete information" and strictly computable, "perfect" solutions to problems. (i.e. tree search, logical inference, planning, etc.)
- ▷ **AI-2** will focus on *probabilistic* scenarios by introducing uncertain situations, and *approximate* solutions to problems. (Bayesian networks, Markov models, machine learning, etc.)
- ▷ **Weak Prerequisites for AI-2:** (if you do not have them, study up as needed)
  - ▷ AI-1 (in particular: PEAS, propositional logic/first-order logic (mostly the syntax), some logic programming)
  - ▷ (very) elementary complexity theory. (big Oh and friends)
  - ▷ rudimentary probability theory (e.g. from stochastics)
  - ▷ basic linear algebra (vectors, matrices,...)
  - ▷ basic real analysis (aka. calculus) (primarily: (partial) derivatives)
- ▷ **Meaning:** I will *assume* you know these things, but some of them we will recap, and what you don't know will make things slightly harder for you, but by no means prohibitively difficult.

"Strict" Prerequisites

- **Most crucially – Mathematical Literacy:** Mathematics is the language that computer scientists express their ideas in! ("*A search problem is a tuple* $(N, S, G, ...)$ *such that...*")
- **Note:** This is a skill that can be *learned*, and more importantly, *practiced!* Not having/honing this skill *will* make things more difficult for you. Be aware of this and, if necessary, work on it – it will pay off, not only in this course.
- **But also:** Motivation, interest, curiosity, hard work. (AI-2 is non-trivial)
- **Note:** Grades correlate significantly with invested effort; including, but not limited to:
  - time spent on exercises, (learning is 80% perspiration, only 20% inspiration)
  - being here in presence, (humans are social animals $\leadsto$ mirror neurons)
  - asking questions, (Q/A dialogues activate brains)
  - talking to your peers, (pool your insights, share your triumphs/frustrations)...

  All of these we try to support with the ALeA system.(which also gives us the data to prove this)

Now we come to a topic that is always interesting to the students: the grading scheme.

Assessment, Grades

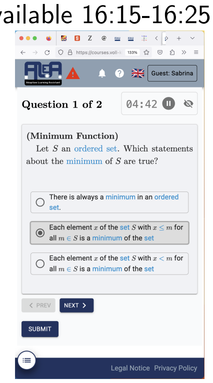
- **Overall (Module) Grade:**
  - Grade via the exam (Klausur) $\leadsto$ 100% of the grade.
  - Up to 10% bonus on-top for an exam with $\geq 50\%$ points. ($< 50\%$ $\leadsto$ no bonus)
  - Bonus points $\widehat{=}$ percentage sum of the best 10 prepquizzes divided by 100.
- **Exam:** exam conducted in presence on paper! ($\sim$ Oct. 10. 2025)
- **Retake Exam:** 90 minutes exam six months later. ($\sim$ April 10. 2026)
- ⚠ You have to register for exams in `https://campo.fau.de` in the first month of classes.
- **Note:** You can de-register from an exam on `https://campo.fau.de` up to three working days before exam. (do not miss that if you are not prepared)

Preparedness Quizzes

- **PrepQuizzes:** Before every lecture we offer a 10 min online quiz – the PrepQuiz – about the material from the previous week. (16:15-16:25; starts in week 2)
- **Motivations:** We do this to
  - keep you prepared and working continuously. (primary)

  - ▷ bonus points if the exam has $\geq 50\%$ points (potential part of your grade)
  - ▷ update the ALeA learner model. (fringe benefit)

- ▷ The prepquizzes will be given in the ALeA system
  - ▷ `https://courses.voll-ki.fau.de/quiz-dash/ai-2`
  - ▷ You have to be logged into ALeA! (via FAU IDM)
  - ▷ You can take the prepquiz on your laptop or phone, ...
  - ▷ ... in the lecture or at home ...
  - ▷ ... via WLAN or 4G Network. (do not overload)
  - ▷ Prepquizzes will only be available 16:15-16:25!



Due to the current AI hype, the course Artificial Intelligence is very popular and thus many degree programs at FAU have adopted it for their curricula. Sometimes the course setup that fits for the CS program does not fit the other's very well, therefore there are some special conditions. I want to state here.

Special Admin Conditions

- ▷ Some degree programs do not "import" the course Artificial Intelligence 1, and thus you may not be able to register for the exam via `https://campo.fau.de`.
  - ▷ Just send me an e-mail and come to the exam, (we do the necessary admin)
  - ▷ Tell your program coordinator about AI-1/2 so that they remedy this situation
- ▷ In "Wirtschafts-Informatik" you can only take AI-1 and AI-2 together in the "Wahlpflichtbereich".
  - ▷ ECTS credits need to be divisible by five $\rightsquigarrow 7.5 + 7.5 = 15$.

I can only warn of what I am aware, so if your degree program lets you jump through extra hoops, please tell me and then I can mention them here.

21.1.2 Getting Most out of AI-2

In this subsection we will discuss a couple of measures that students may want to consider to get most out of the AI-2 course.

None of the things discussed in this subsection – homeworks, tutorials, study groups, and attendance – are mandatory (we cannot force you to do them; we offer them to you as learning opportunities), but most of them are very clearly correlated with success (i.e. passing the exam and getting a good grade), so taking advantage of them may be in your own interest.

AI-2 Homework Assignments

- **Goal:** Homework assignments reinforce what was taught in lectures.
- **Homework Assignments:** Small individual problem/programming/proof task
  - but take time to solve (at least read them directly ⇝ questions)
- **Didactic Intuition:** Homework assignments give you material to test your understanding and show you how to apply it.
- ⚠ Homeworks give no points, but without trying you are unlikely to pass the exam.
- **Our Experience:** Doing your homework is probably even *more* important (and predictive of exam success) than attending the lecture in person!
- Homeworks will be mainly peer-graded in the ALeA system.
- **Didactic Motivation:** Through peer grading students are able to see mistakes in their thinking and can correct any problems in future assignments. By grading assignments, students may learn how to complete assignments more accurately and how to improve their future results. (not just us being lazy)

It is very well-established experience that without doing the homework assignments (or something similar) on your own, you will not master the concepts, you will not even be able to ask sensible questions, and take very little home from the course. Just sitting in the course and nodding is not enough!

AI-2 Homework Assignments – Howto

- **Homework Workflow:** in ALeA (see below)
  - Homework assignments will be published on thursdays: see `https://courses.voll-ki.fau.de/hw/ai-1`
  - Submission of solutions via the ALeA system in the week after
  - Peer grading/feedback (and master solutions) via answer classes.
- **Quality Control:** TAs and instructors will monitor and supervise peer grading.
- **Experiment:** Can we motivate enough of you to make peer assessment self-sustaining?
  - I am appealing to your sense of community responsibility here …
  - You should only expect other's to grade your submission if you grade their's (cf. Kant's "Moral Imperative")
  - **Make no mistake**: The grader usually learns at least as much as the gradee.
- **Homework/Tutorial Discipline:**

- ▷ Start early! (many assignments need more than one evening's work)
- ▷ Don't start by sitting at a blank screen (talking & study groups help)
- ▷ Humans will be trying to understand the text/code/math when grading it.
- ▷ Go to the tutorials, discuss with your TA! (they are there for you!)

If you have questions please make sure you discuss them with the instructor, the teaching assistants, or your fellow students. There are three sensible venues for such discussions: online in the lectures, in the tutorials, which we discuss now, or in the course forum – see below. Finally, it is always a very good idea to form study groups with your friends.

Tutorials for Artificial Intelligence 1

- ▷ **Approach:** Weekly tutorials and homework assignments (first one in week two)
- ▷ **Goal 1:** Reinforce what was taught in the lectures. (you need practice)
- ▷ **Goal 2:** Allow you to ask any question you have in a protected environment.
- ▷ **Instructor/Lead TA:** Florian Rabe (KWARC Postdoc, Privatdozent)
  - ▷ Room: 11.137 @ Händler building, `florian.rabe@fau.de`
- ▷ **Tutorials:** One each taught by Florian Rabe (lead); Primula Mukherjee, Ilhaam Shaikh, Praveen Kumar Vadlamani, and Shreya Rajesh More.
  - ▷ Tutorials will start in week 3. (before there is nothing to do)
  - ▷ Details (rooms, times, etc) will be announced in time (i.e. not now) on the forum and matrix channel.
- ▷ **Life-saving Advice:** Go to your tutorial, and prepare for it by having looked at the slides and the homework assignments!

Collaboration

- ▷ **Definition 21.1.1.** Collaboration (or cooperation) is the process of groups of agents acting together for common, mutual benefit, as opposed to acting in competition for selfish benefit. In a collaboration, every agent contributes to the common goal and benefits from the contributions of others.
- ▷ In learning situations, the benefit is "better learning".
- ▷ **Observation:** In collaborative learning, the overall result can be significantly better than in competitive learning.
- ▷ **Good Practice:** Form study groups. (long- or short-term)
  1. ⚠ Those learners who work/help most, learn most!
  2. ⚠ Freeloaders – individuals who only watch – learn very little!

▷ It is OK to collaborate on homework assignments in AI-2! (no bonus points)

▷ Choose your study group well! (ALeA helps via the study buddy feature)

As we said above, almost all of the components of the AI-2 course are optional. That even applies to attendance. But make no mistake, attendance is important to most of you. Let me explain, . . .

Do I need to attend the AI-2 Lectures

▷ Attendance is not mandatory for the AI-2 course. (official version)

▷ **Note:** There are two ways of learning: (both are OK, your mileage may vary)

- ▷ Approach B: Read a book/papers (here: lecture notes)
- ▷ Approach I: come to the lectures, be involved, interrupt the instructor whenever you have a question.

The only advantage of I over B is that books/papers do not answer questions

▷ Approach S: come to the lectures and sleep does not work!

▷ The closer you get to research, the more we need to discuss!

Do use the opportunity to discuss the AI-2 topics with others. After all, one of the non-trivial skills you want to learn in the course is how to talk about artificial intelligence topics. And that takes practice, practice, and practice.

21.1.3 Learning Resources for AI-2

Textbooks and supplementary Literature

▷ **Textbook:** *Russel/Norvig: Artificial Intelligence, A modern Approach* [**RusNor:AIMA09**].

- ▷ basically "broad but somewhat shallow"
- ▷ great to get intuitions on the basics of AI

Make sure that you read the edition ≥ 3 $\leadsto$ vastly improved over ≤ 2.

Course Notes, Forum, Matrix

▷ **Lecture notes** will be posted at `https://kwarc.info/teaching/AI`

- ▷ We mostly prepare/update them as we go along (semantically preloaded $\leadsto$ research resource)
- ▷ Please report any errors/shortcomings you notice. (improve for the group/successors)

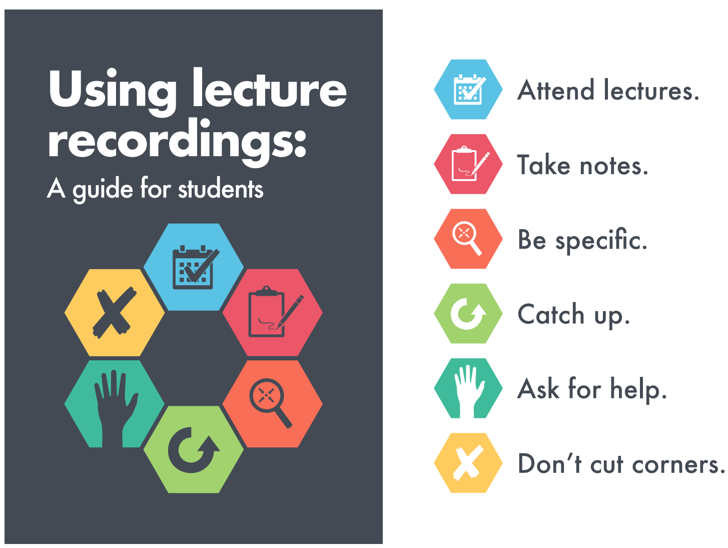
- ▷ **StudOn Forum:** For announcements – `https://www.studon.fau.de/studon/goto.php?target=lcode_70Bjcaxg`
- ▷ **Matrix Channel:** `https://matrix.to/#/#ai-12:fau.de` for questions, discussion with instructors and among your fellow students. (your channel, use it!)

  Login via FAU IDM ⇝ instructions
- ▷ **Course Videos** are at at `https://fau.tv/course/id/4225`.
- ▷ **Do not let the videos mislead you:** Coming to class is highly correlated with passing the exam!

FAU has issued a very insightful guide on using lecture videos. It is a good idea to heed these recommendations, even if they seem annoying at first.

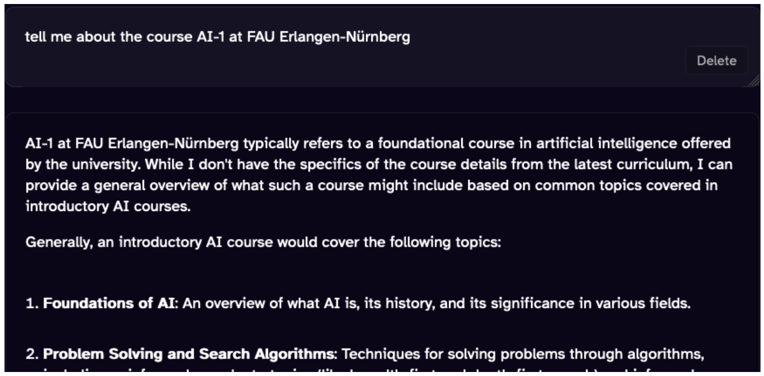
Practical recommendations on Lecture Videos

- ▷ **Excellent Guide:** [**NorKueRob:lcprs18**] (German version at [**NorKueRob:vnas18**])



NOT a Resource for : LLMs – AI-based tools like ChatGPT

- ▷ **Definition 21.1.2.** A large language model (LLM) is a computational model capable of language generation or other natural language processing tasks.
- ▷ **Example 21.1.3.** OpenAI's GPT, Google's Bard, and Meta's Llama.
- ▷ **Definition 21.1.4.** A chatbot is a software application or web interface that is designed to mimic human conversation through text or voice interactions. Modern chatbots are usually based on LLMs.
- ▷ **Example 21.1.5 (ChatGPT talks about AI-1).** (but remains vague)

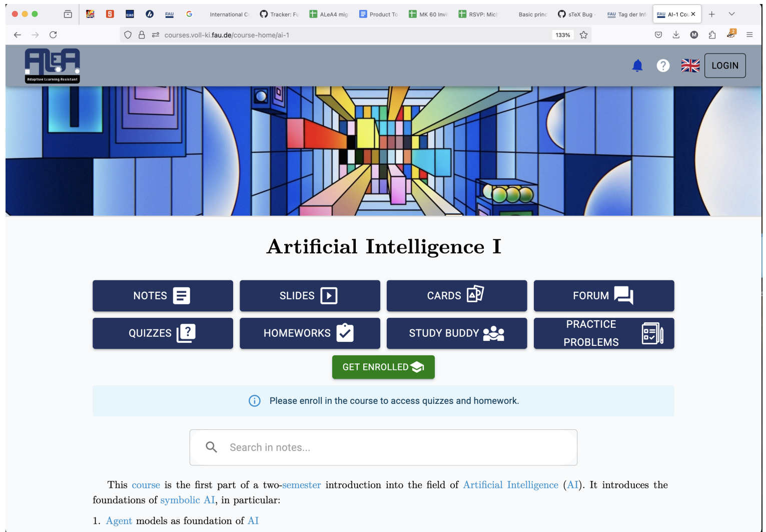


- ▷ **Note:** LLM-based chatbots invent *every word*! (surprisingly often correct)
- ▷ **Example 21.1.6 (In the AI-1 exam).** ChatGPT scores ca. 50% of the points.
  - ▷ ChatGPT can almost pass the exam . . . (We could award it a Master's degree)
  - ▷ But can you? (the AI-1 exams will be in person on paper)

  You will only pass the exam, if you can do AI-1 yourself!
- ▷ **Intuition:** AI tools like GhatGPT, CoPilot, etc. (see also [Shein:iacse24])
  - ▷ can help you solve problems, (valuable tools in production situations)
  - ▷ hinders learning if used for homeworks/quizzes, etc. (like driving instead of jogging)
- ▷ **What (not) to do:** (to get most of the brave new AI-supported world)
  - ▷ try out these tools to get a first-hand intuition what they can/cannot do
  - ▷ challenge yourself while learning so that you can also do it (mind over matter!)

ALEA in AI-2

- ▷ We assume that you already know the ALEA system from last semester



▷ Use it for

- ▷ lecture notes (notes- vs slides-oriented)
- ▷ flashcards (drill yourself on the AI-2 jargon/concepts)
- ▷ course forum (questions, discussions and error reporting)
- ▷ solving and peer-grading homework assignments
- ▷ finding study groups (you need not endure AI-2 alone)
- ▷ practicing with targeted problems (e.g. from old exams)
- ▷ doing the prepquizzes (before each lecture)

Excursion: We will recap an introduction to ALeA system in???.

21.2 Overview over AI and Topics of AI-II

We restart the new semester by reminding ourselves of (the problems, methods, and issues of) artificial intelligence, and what has been achived so far.

21.2.1 What is Artificial Intelligence?

The first question we have to ask ourselves is "What is artificial intelligence?", i.e. how can we define it. And already that poses a problem since the natural definition *like human intelligence, but artificially realized* presupposes a definition of intelligence, which is equally problematic; even Psychologists and Philosophers – the subjects nominally "in charge" of natural intelligence – have problems defining it, as witnessed by the plethora of theories e.g. found at [**wiki:human_intelligence**].

What is Artificial Intelligence? Definition

▷ **Definition 21.2.1 (According to Wikipedia).** Artificial Intelligence (AI) is intelligence exhibited by machines

▷ **Definition 21.2.2 (also).** Artificial Intelligence (AI) is a sub-field of CS that is concerned with the automation of intelligent behavior.

▷ **BUT:** it is already difficult to define intelligence precisely.

▷ **Definition 21.2.3 (Elaine Rich).** artificial intelligence (AI) studies how we can make the computer do things that humans can still do better at the moment.

Maybe we can get around the problems of defining "what artificial intelligence is", by just describing the necessary components of AI (and how they interact). Let's have a try to see whether that is more informative.

What is Artificial Intelligence? Components

▷ **Elaine Rich:** AI studies how we can make the computer do things that humans can still do better at the moment.

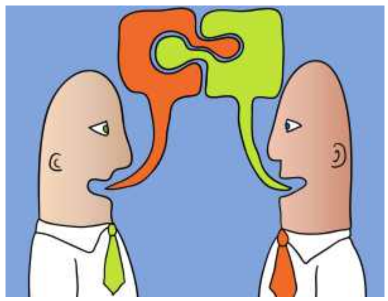
▷ This needs a combination of

the ability to learn

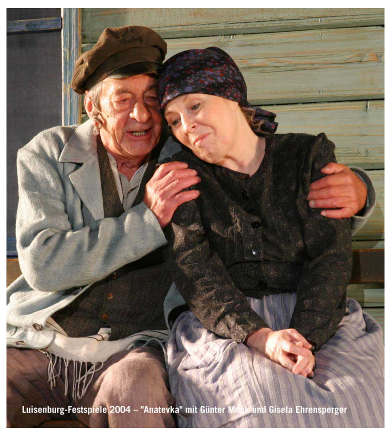
Inference

Perception

Language understanding



Emotion

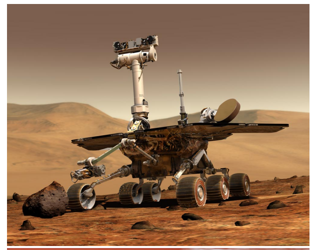


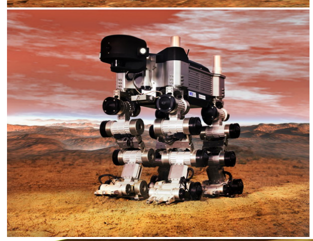
Note that list of components is controversial as well. Some say that it lumps together cognitive capacities that should be distinguished or forgets others, ... We state it here much more to get AI-2 students to think about the issues than to make it normative.

21.2.2 Artificial Intelligence is here today!

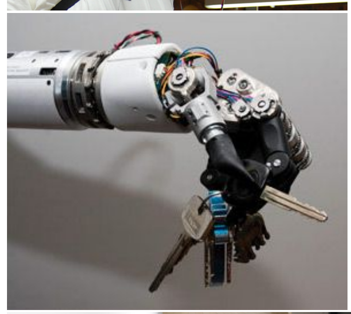
The components of artificial intelligence are quite daunting, and none of them are fully understood, much less achieved artificially. But for some tasks we can get by with much less. And indeed that is what the field of artificial intelligence does in practice – but keeps the lofty ideal around. This practice of "trying to achieve AI in selected and restricted domains" (cf. the discussion starting with slide **??**) has borne rich fruits: systems that meet or exceed human capabilities in such areas. Such systems are in common use in many domains of application.

Artificial Intelligence is here today!



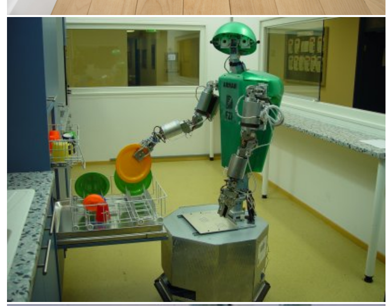


- in outer space
  - in outer space systems need autonomous control:
  - remote control impossible due to time lag



- in artificial limbs
  - the user controls the prosthesis via existing nerves, can e.g. grip a sheet of paper.

- in household appliances
  - The iRobot Roomba vacuums, mops, and sweeps in corners, . . . , parks, charges, and discharges.



  - general robotic household help is on the horizon.
- in hospitals

  - in the USA 90% of the prostate operations are carried out by RoboDoc
  - Paro is a cuddly robot that eases solitude in nursing homes.

FAU : 709 2025-05-01

We will conclude this subsection with a note of caution.

The AI Conundrum

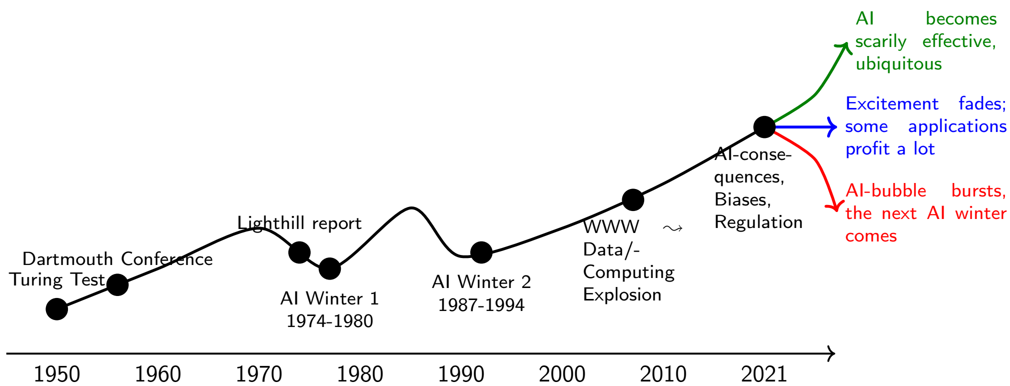
- **Observation:** Reserving the term "artificial intelligence" has been quite a land grab!
- **But:** researchers at the Dartmouth Conference (1956) really thought they would solve/reach AI in two/three decades.
- **Consequence:** AI still asks the big questions. (and still promises answers soon)
- **Another Consequence:** AI as a field is an incubator for many innovative technologies.
- **AI Conundrum:** Once AI solves a subfield it is called "CS".(becomes a separate subfield of CS)
- **Example 21.2.4.** Functional/Logic Programming, automated theorem proving, Planning, machine learning, Knowledge Representation, …
- **Still Consequence:** AI research was alternatingly flooded with money and cut off brutally.

FAU : 710 2025-05-01

All of these phenomena can be seen in the growth of AI as an academic discipline over the course of its now over 70 year long history.

The current AI Hype — Part of a longer Story

- The history of AI as a discipline has been very much tied to the amount of funding – that allows us to do research and development.
- Funding levels are tied to public perception of success (especially for AI)
- **Definition 21.2.5.** An AI winter is a time period of low public perception and funding for AI,
  mostly because AI has failed to deliver on its – sometimes overblown – promises
  An AI summer is a time period of high public perception and funding for AI
- A potted history of AI (AI summers and summers)



Of course, the future of AI is still unclear, we are currently in a massive hype caused by the advent of deep neural networks being trained on all the data of the Internet, using the computational power of huge compute farms owned by an oligopoly of massive technology companies – we are definitely in an AI summer.

But AI as a academic community and the tech industry also make outrageous promises, and the media pick it up and distort it out of proportion, . . . So public opinion could flip again, sending AI into the next winter.

21.2.3 Ways to Attack the AI Problem

There are currently three main avenues of attack to the problem of building artificially intelligent systems. The (historically) first is based on the symbolic representation of knowledge about the world and uses inference-based methods to derive new knowledge on which to base action decisions. The second uses statistical methods to deal with uncertainty about the world state and learning methods to derive new (uncertain) world assumptions to act on.

Four Main Approaches to Artificial Intelligence

- ▷ **Definition 21.2.6.** Symbolic AI is a subfield of AI based on the assumption that many aspects of intelligence can be achieved by the manipulation of symbols, combining them into meaning-carrying structures (expressions) and manipulating them (using processes) to produce new expressions.
- ▷ **Definition 21.2.7.** Statistical AI remedies the two shortcomings of symbolic AI approaches: that all concepts represented by symbols are crisply defined, and that all aspects of the world are knowable/representable in principle. Statistical AI adopts sophisticated mathematical models of uncertainty and uses them to create more accurate world models and reason about them.
- ▷ **Definition 21.2.8.** Subsymbolic AI (also called connectionism or neural AI) is a subfield of AI that posits that intelligence is inherently tied to brains, where information is represented by a simple sequence pulses that are processed in parallel via simple calculations realized by neurons, and thus concentrates on neural computing.
- ▷ **Definition 21.2.9.** Embodied AI posits that intelligence cannot be achieved by reasoning about the state of the world (symbolically, statistically, or connectivist), but must be embodied i.e. situated in the world, equipped with a "body" that can interact with it via sensors and actuators. Here, the main method for realizing intelligent behavior is by learning from the world.

As a consequence, the field of artificial intelligence (AI) is an engineering field at the intersection of CS (logic, programming, applied statistics), Cognitive Science (psychology, neuroscience), philosophy (can machines think, what does that mean?), linguistics (natural language understanding), and mechatronics (robot hardware, sensors).

Subsymbolic AI and in particular machine learning is currently hyped to such an extent, that many people take it to be synonymous with "Artificial Intelligence". It is one of the goals of this course to show students that this is a very impoverished view.

Two ways of reaching Artificial Intelligence?

▷ We can classify the AI approaches by their coverage and the analysis depth (they are complementary)

Deep	symbolic AI-1	not there yet cooperation?
Shallow	no-one wants this	statistical/sub symbolic AI-2
Analysis ↑ vs. Coverage →	Narrow	Wide

▷ **This semester** we will cover foundational aspects of symbolic AI (deep/narrow processing)

▷ **next semester** concentrate on statistical/subsymbolic AI. (shallow/wide-coverage)

We combine the topics in this way in this course, not only because this reproduces the historical development but also as the methods of statistical and subsymbolic AI share a common basis.
It is important to notice that all approaches to AI have their application domains and strong points. We will now see that exactly the two areas, where symbolic AI and statistical/subsymbolic AI have their respective fortes correspond to natural application areas.

Environmental Niches for both Approaches to AI

▷ **Observation:** There are two kinds of applications/tasks in AI

- ▷ Consumer tasks: consumer grade applications have tasks that must be fully generic and wide coverage. (e.g. machine translation like Google Translate)
- ▷ Producer tasks: producer grade applications must be high-precision, but can be domain-specific (e.g. multilingual documentation, machinery-control, program verification, medical technology)

Precision			
100%	Producer Tasks		
50%		Consumer Tasks	
	$10^{3\pm1}$ Concepts	$10^{6\pm1}$ Concepts	**Coverage**

after Aarne Ranta **[Ranta:atcp17]**.

▷ **General Rule:** Subsymbolic AI is well suited for consumer tasks, while symbolic AI is better suited for producer tasks.

▷ A domain of producer tasks I am interested in: mathematical/technical documents.

An example of a producer task – indeed this is where the name comes from – is the case of a machine tool manufacturer T, which produces digitally programmed machine tools worth multiple million Euro and sells them into dozens of countries. Thus T must also provide comprehensive machine operation manuals, a non-trivial undertaking, since no two machines are identical and they must be translated into many languages, leading to hundreds of documents. As those manual share a lot of semantic content, their management should be supported by AI techniques. It is critical that these methods maintain a high precision, operation errors can easily lead to very costly machine damage and loss of production. On the other hand, the domain of these manuals is quite restricted. A machine tool has a couple of hundred components only that can be described by a couple of thousand attributes only.

Indeed companies like T employ high-precision AI techniques like the ones we will cover in this course successfully; they are just not so much in the public eye as the consumer tasks.

21.2.4 AI in the KWARC Group

The KWARC Research Group

- **Observation:** The ability to represent knowledge about the world and to draw logical inferences is one of the central components of intelligent behavior.
- **Thus:** reasoning components of some form are at the heart of many AI systems.
- **KWARC Angle:** Scaling up (web-coverage) without dumbing down (too much)
  - Content markup instead of full formalization (too tedious)
  - User support and quality control instead of "The Truth" (elusive anyway)
  - use Mathematics as a test tube (⚠ Mathematics $\hat{=}$ Anything Formal ⚠)
  - care more about applications than about philosophy (we cannot help getting this right anyway as logicians)
- The KWARC group was established at Jacobs Univ. in 2004, moved to FAU Erlangen in 2016
- See `http://kwarc.info` for projects, publications, and links

Overview: KWARC Research and Projects

Applications: eMath 3.0, Active Documents, Active Learning, Semantic Spreadsheets/CAD/CAM, Change Mangagement, Global Digital Math Library, Math Search Systems, SMGloM: Semantic Multilingual Math Glossary, Serious Games, ...		
Foundations of Math: ▷ MathML, *OpenMath* ▷ advanced Type Theories ▷ MMT: Meta Meta Theory ▷ Logic Morphisms/Atlas ▷ Theorem Prover/CAS Interoperability ▷ Mathematical Models/Simulation	**KM & Interaction**: ▷ Semantic Interpretation (aka. Framing) ▷ math-literate interaction ▷ MathHub: math archives & active docs ▷ Active documents: embedded semantic services ▷ Model-based Education	**Semantization**: ▷ LaTeXML: LaTeX ⇝ XML ▷ sTeX: Semantic LaTeX ▷ invasive editors ▷ Context-Aware IDEs ▷ Mathematical Corpora ▷ Linguistics of Math ▷ ML for Math Semantics Extraction
Foundations: Computational Logic, Web Technologies, OMDoc/MMT		

Research Topics in the KWARC Group

- ▷ We are always looking for bright, motivated KWARCies.
- ▷ We have topics in for all levels! (Enthusiast, Bachelor, Master, Ph.D.)
- ▷ List of current topics: `https://gl.kwarc.info/kwarc/thesis-projects/`
  - ▷ Automated Reasoning: Maths Representation in the Large
  - ▷ Logics development, (Meta)^n-Frameworks
  - ▷ Math Corpus Linguistics: Semantics Extraction
  - ▷ Serious Games, Cognitive Engineering, Math Information Retrieval, Legal Reasoning, ...
  - ▷ ... last but not least: KWARC is the home of ALeA!
- ▷ We always try to find a topic at the intersection of your and our interests.
- ▷ We also sometimes have positions!. (HiWi, Ph.D.: $\frac{1}{2}$ E-13, PostDoc: full E-13)

21.2.5 Agents and Environments in AI2

This part of the lecture notes addresses inference and agent decision making in partially observable environments, i.e. where we only know probabilities instead of certainties whether propositions are true/false. We cover basic probability theory and – based on that – Bayesian Networks and simple decision making in such environments. Finally we extend this to probabilistic temporal models and their decision theory.

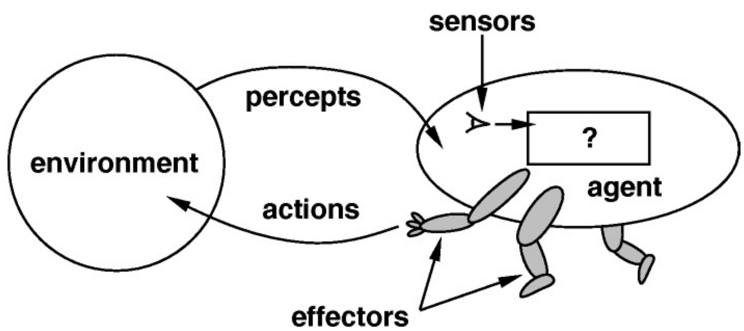
21.2.5.1 Recap: Rational Agents as a Conceptual Framework

Agents and Environments

▷ **Definition 21.2.10.** An agent is anything that

- ▷ perceives its environment via sensors (a means of sensing the environment)
- ▷ acts on it with actuators (means of changing the environment).

Any recognizable, coherent employment of the actuators of an agent is called an action.



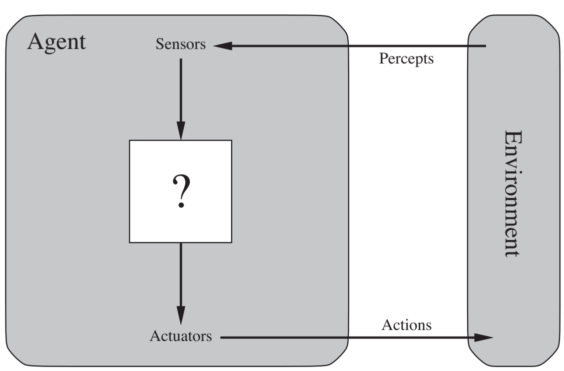
▷ **Example 21.2.11.** Agents include humans, robots, softbots, thermostats, etc.

▷ **Remark:** The notion of an agent and its environment is intentionally designed to be inclusive. We will classify and discuss subclasses of both later.

One possible objection to this is that the agent and the environment are conceptualized as separate entities; in particular, that the image suggests that the agent itself is not part of the environment. Indeed that is intended, since it makes thinking about agents and environments easier and is of little consequence in practice. In particular, the offending separation is relatively easily fixed if needed.

Agent Schema: Visualizing the Internal Agent Structure

▷ **Agent Schema:** We will use the following kind of agent schema to visualize the internal structure of an agent:



Different agents differ on the contents of the white box in the center.

Rationality

- **Idea:** Try to design agents that are successful! (aka. "do the right thing")
- **Problem:** What do we mean by "successful", how do we measure "success"?
- **Definition 21.2.12.** A performance measure is a function that evaluates a sequence of environments.
- **Example 21.2.13.** A performance measure for a vacuum cleaner could
  - award one point per "square" cleaned up in time T?
  - award one point per clean "square" per time step, minus one per move?
  - penalize for $> k$ dirty squares?
- **Definition 21.2.14.** An agent is called rational, if it chooses whichever action maximizes the expected value of the performance measure given the percept sequence to date.
- **Critical Observation:** We only need to maximize the expected value, not the actual value of the performance measure!
- **Question:** Why is rationality a good quality to aim for?

Let us see how the observation that we only need to maximize the expected value, not the actual value of the performance measure affects the consequences.

Consequences of Rationality: Exploration, Learning, Autonomy

- **Note:** A rational agent need not be perfect:
  - It only needs to maximize expected value (rational $\neq$ omniscient)
    - need not predict e.g. very unlikely but catastrophic events in the future
  - Percepts may not supply all relevant information (rational $\neq$ clairvoyant)
    - if we cannot perceive things we do not need to react to them.
    - but we may need to try to find out about hidden dangers (exploration)
  - Action outcomes may not be as expected (rational $\neq$ successful)
    - but we may need to take action to ensure that they do (more often) (learning)
- **Note:** Rationality may entail exploration, learning, autonomy (depending on the environment / task)
- **Definition 21.2.15.** An agent is called autonomous, if it does not rely on the prior knowledge about the environment of the designer.
- Autonomy avoids fixed behaviors that can become unsuccessful in a changing environment. (anything else would be irrational)
- The agent may have to learn all relevant traits, invariants, properties of the environment and actions.

For the design of agent for a specific task – i.e. choose an agent architecture and design an agent program, we have to take into account the performance measure, the environment, and the characteristics of the agent itself; in particular its actions and sensors.

PEAS: Describing the Task Environment

- **Observation:** To design a rational agent, we must specify the task environment in terms of performance measure, environment, actuators, and sensors, together called the PEAS components.
- **Example 21.2.16.** When designing an automated taxi:
  - **Performance measure:** safety, destination, profits, legality, comfort, ...
  - **Environment:** US streets/freeways, traffic, pedestrians, weather, ...
  - **Actuators:** steering, accelerator, brake, horn, speaker/display, ...
  - **Sensors:** video, accelerometers, gauges, engine sensors, keyboard, GPS, ...
- **Example 21.2.17 (Internet Shopping Agent).** The task environment:
  - Performance measure: price, quality, appropriateness, efficiency
  - Environment: current and future WWW sites, vendors, shippers
  - Actuators: display to user, follow URL, fill in form
  - Sensors: HTML pages (text, graphics, scripts)

The PEAS criteria are essentially a laundry list of what an agent design task description should include.

Environment types

- **Observation 21.2.18.** *Agent design is largely determined by the type of environment it is intended for.*
- **Problem:** There is a vast number of possible kinds of environments in AI.
- **Solution:** Classify along a few "dimensions". (independent characteristics)
- **Definition 21.2.19.** For an agent a we classify the environment e of a by its type, which is one of the following. We call e

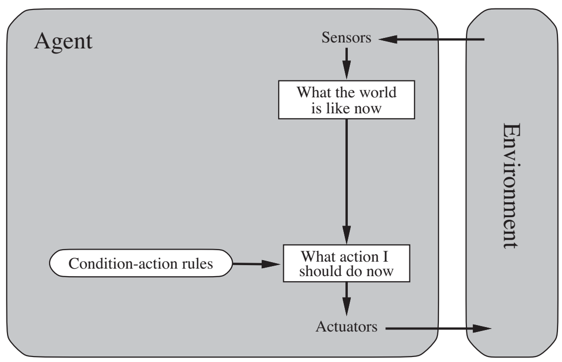
  1. fully observable, iff the a's sensors give it access to the complete state of the environment at any point in time, else partially observable.
  2. deterministic, iff the next state of the environment is completely determined by the current state and a's action, else stochastic.
  3. episodic, iff a's experience is divided into atomic episodes, where it perceives and then performs a single action. Crucially, the next episode does not depend on previous ones. Non-episodic environments are called sequential.
  4. dynamic, iff the environment can change without an action performed by a, else static. If the environment does not change but a's performance measure does, we call e semidynamic.
  5. discrete, iff the sets of e's state and a's actions are countable, else continuous.

6. single-agent, iff only a acts on e; else multi-agent (when must we count parts of e as agents?)

Reflex Agents

▷ **Definition 21.2.20.** An agent $\langle\mathcal{P},\mathcal{A},f\rangle$ is called a reflex agent, iff it only takes the last percept into account when choosing an action, i .e. $f(p_1,\ldots,p_k)=f(p_k)$ for all $p_1,\ldots,p_k\in\mathcal{P}$.

▷ **Agent Schema:**



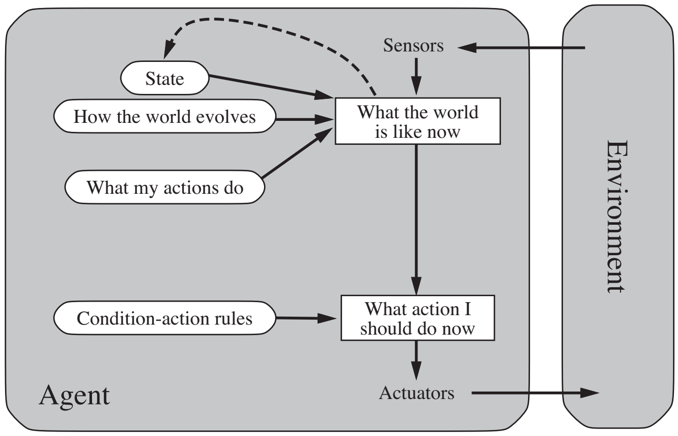
▷ **Example 21.2.21 (Agent Program).**

```
procedure Reflex−Vacuum−Agent [location,status] returns an action
  if status = Dirty then . . .
```


Model-based Reflex Agents: Idea

▷ **Idea:** Keep track of the state of the world we cannot see in an internal model.

▷ **Agent Schema:**



Model-based Reflex Agents: Definition

▷ **Definition 21.2.22.** A model-based agent $\langle\mathcal{P},\mathcal{A},\mathcal{S},\mathcal{T},s_0,S,a\rangle$ is an agent $\langle\mathcal{P},\mathcal{A},f\rangle$ whose actions depend on

1. a world model: a set $\mathcal{S}$ of possible states, and a start state $s_0\in\mathcal{S}$.
2. a transition model $\mathcal{T}$, that predicts a new state $\mathcal{T}(s,a)$ from a state s and an action a.
3. a sensor model S that given a state s and a percept p determine a new state $S(s,p)$.
4. an action function $a\colon\mathcal{S}\to\mathcal{A}$ that given a state selects the next action.

If the world model of a model-based agent A is in state s and A has last taken action a, and now perceives p, then A will transition to state $s'=S(p,\mathcal{T}(s,a))$ and take action $a'=a(s')$.

So, given a sequence $p_1,\ldots,p_n$ of percepts, we recursively define states $s_n=S(\mathcal{T}(s_{n-1},a(s_{n-1})),p_n)$ with $s_1=S(s_0,p_1)$. Then $f(p_1,\ldots,p_n)=a(s_n)$.

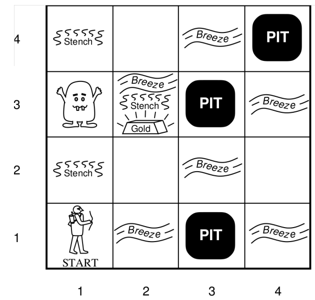
▷ **Note:** As different percept sequences lead to different states, so the agent function $f()\colon\mathcal{P}^*\to\mathcal{A}$ no longer depends only on the last percept.

▷ **Example 21.2.23 (Tail Lights Again).** Model-based agents can do the ??? if the states include a concept of tail light brightness.

21.2.5.2 Sources of Uncertainty

Sources of Uncertainty in Decision-Making

Where's that d. . . Wumpus?
And where am I, anyway??



- **Non-deterministic actions:**
  - "When I try to go forward in this dark cave, I might actually go forward-left or forward-right."
- **Partial observability with unreliable sensors:**
  - "Did I feel a breeze right now?";
  - "I think I might smell a Wumpus here, but I got a cold and my nose is blocked."
  - "According to the heat scanner, the Wumpus is probably in cell [2,3]."
- **Uncertainty about the domain behavior:**
  - "Are you *sure* the Wumpus never moves?"

Unreliable Sensors

- **Robot Localization:** Suppose we want to support localization using landmarks to narrow down the area.
- **Example 21.2.24.** *If you see the Eiffel tower, then you're in Paris.*
- **Difficulty:** Sensors can be imprecise.
  - Even if a landmark is perceived, we cannot conclude with certainty that the robot is at that location.
  - *This is the half-scale Las Vegas copy, you dummy.*
  - Even if a landmark is *not* perceived, we cannot conclude with certainty that the robot is *not* at that location.
  - *Top of Eiffel tower hidden in the clouds.*
- Only the probability of being at a location increases or decreases.

21.2.5.3 Agent Architectures based on Belief States

We are now ready to proceed to environments which can only partially observed and where actions are non deterministic. Both sources of uncertainty conspire to allow us only partial knowledge about the world, so that we can only optimize "expected utility" instead of "actual utility" of our actions.

World Models for Uncertainty

- **Problem:** We do not know with certainty what state the world is in!
- **Idea:** Just keep track of all the possible states it could be in.
- **Definition 21.2.25.** A model-based agent has a world model consisting of
  - a belief state that has information about the possible states the world may be in,
  - a sensor model that updates the belief state based on sensor information, and
  - a transition model that updates the belief state based on actions.
- **Idea:** The agent environment determines what the world model can be.
- In a fully observable, deterministic environment,
  - we can observe the initial state and subsequent states are given by the actions alone.
  - Thus the belief state is a singleton (we call its sole member the world state) and the transition model is a function from states and actions to states: a transition function.

That is exactly what we have been doing until now: we have been studying methods that build on descriptions of the "actual" world, and have been concentrating on the progression from atomic to factored and ultimately structured representations. Tellingly, we spoke of "world states" instead of "belief states"; we have now justified this practice in the brave new belief-based world models by the (re-) definition of "world states" above. To fortify our intuitions, let us recap from a belief-state-model perspective.

World Models by Agent Type in AI-1

- **Search-based Agents:** In a fully observable, deterministic environment
  - goal-based agent with world state $\widehat{=}$ "current state"
  - no inference. (goal $\widehat{=}$ goal state from search problem)
- **CSP-based Agents:** In a fully observable, deterministic environment
  - goal-based agent withworld state $\widehat{=}$ constraint network,
  - inference $\widehat{=}$ constraint propagation. (goal $\widehat{=}$ satisfying assignment)
- **Logic-based Agents:** In a fully observable, deterministic environment
  - model-based agent with world state $\widehat{=}$ logical formula
  - inference $\widehat{=}$ e.g. DPLL or resolution.
- **Planning Agents:** In a fully observable, deterministic, environment
  - goal-based agent with world state $\widehat{=}$ PL0, transition model $\widehat{=}$ STRIPS,
  - inference $\widehat{=}$ state/plan space search. (goal: complete plan/execution)

Let us now see what happens when we lift the restrictions of total observability and determinism.

World Models for Complex Environments

- In a fully observable, but stochastic environment,
  - the belief state must deal with a set of possible states.
  - $\rightsquigarrow$ generalize the transition function to a transition relation.
- **Note:** This even applies to online problem solving, where we can just perceive the state. (e.g. when we want to optimize utility)
- In a deterministic, but partially observable environment,
  - the belief state must deal with a set of possible states.
  - we can use transition functions.
  - We need a sensor model, which predicts the influence of percepts on the belief state – during update.
- In a stochastic, partially observable environment,
  - mix the ideas from the last two. (sensor model + transition relation)

Preview: New World Models (Belief) $\rightsquigarrow$ new Agent Types

- **Probabilistic Agents:** In a partially observable environment
  - belief state $\widehat{=}$ Bayesian networks,
  - inference $\widehat{=}$ probabilistic inference.
- **Decision-Theoretic Agents:** In a partially observable, stochastic environment
  - belief state + transition model $\widehat{=}$ decision networks,
  - inference $\widehat{=}$ maximizing expected utility.
- We will study them in detail this semester.

Overview: AI2

- Basics of probability theory (probability spaces, random variables, conditional probabilities, independence,...)
- Probabilistic reasoning: Computing the *a posteriori* probabilities of events given evidence, causal reasoning (Representing distributions efficiently, Bayesian networks,...)
- Probabilistic Reasoning over time (Markov chains, Hidden Markov models,...)

$\Rightarrow$ We can update our world model episodically based on observations (i.e. sensor data)

▷ Decision theory: Making decisions under uncertainty (Preferences, Utilities, Decision networks, Markov Decision Procedures,...)

⇒ We can choose the right action based on our world model and the likely outcomes of our actions

▷ Machine learning: Learning from data (Decision Trees, Classifiers, Neural Networks,...)

Appendix A

Excursions

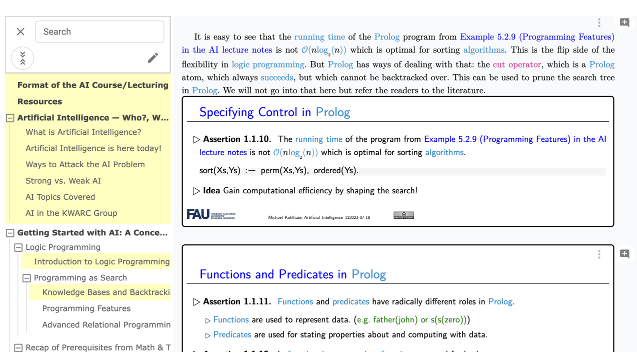
As this course is predominantly an overview over the topics of artificial intelligence, and not about the theoretical underpinnings, we give the discussion about these as a “suggested readings” chapter here.

A.1 ALeA – AI-Supported Learning

In this section we introduce the ALeA (Adaptive Learning Assistant) system, a learning support system we will use to support students in AI-2.

ALeA: Adaptive Learning Assistant

- ▷ **Idea:** Use AI methods to help teach/learn AI (AI4AI)
- ▷ **Concretely:** Provide HTML versions of the AI-2 slides/lecture notes and embed learning support services into them. (for pre/postparation of lectures)
- ▷ **Definition A.1.1.** Call a document active, iff it is interactive and adapts to specific information needs of the readers. (lecture notes on steroids)
- ▷ **Intuition:** ALeA serves active course materials. (PDF mostly inactive)
- ▷ **Goal:** Make ALeA more like a instructor + study group than like a book!
- ▷ **Example A.1.2 (Course Notes).** $\widehat{=}$ Slides + Comments



$\rightsquigarrow$ yellow parts in table of contents (left) already covered in lectures.

The central idea in the AI4AI approach – using AI to support learning AI – and thus the ALeA system is that we want to make course materials – i.e. what we give to students for preparing and postparing lectures – more like teachers and study groups (only available 24/7) than like static books.

VoLL-KI Portal at `https://courses.voll-ki.fau.de`

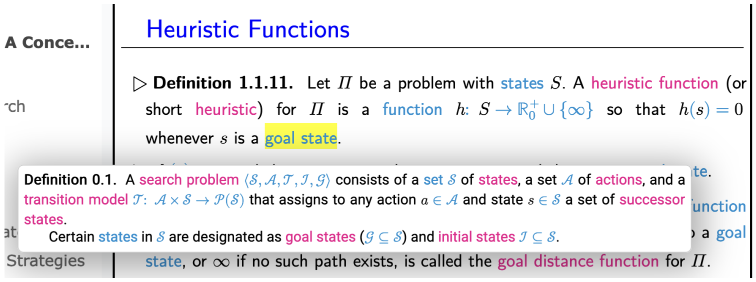
- **Portal for ALeA Courses:** `https://courses.voll-ki.fau.de`

- **AI-2 in ALeA:** `https://courses.voll-ki.fau.de/course-home/ai-2`
  - All details for the course.
  - recorded syllabus (keep track of material covered in course)
  - syllabus of the last semesters (for over/preview)
- **ALeA Status:** The ALEA system is deployed at FAU for over 1000 students taking eight courses
  - (some) students use the system actively (our logs tell us)
  - reviews are mostly positive/enthusiastic (error reports pour in)

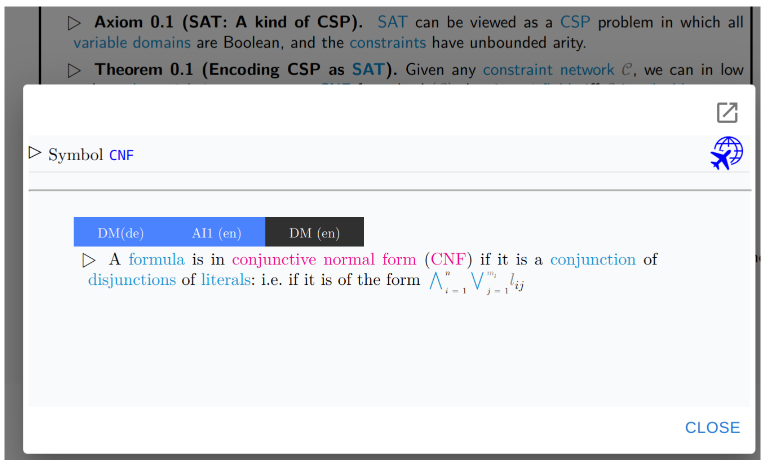
The ALEA AI-2 page is the central entry point for working with the ALeA system. You can get to all the components of the system, including two presentations of the course contents (notes- and slides-centric ones), the flashcards, the localized forum, and the quiz dashboard.
We now come to the heart of the ALeA system: its learning support services, which we will now briefly introduce. Note that this presentation is not really sufficient to undertstand what you may be getting out of them, you will have to try them, and interact with them sufficiently that the learner model can get a good estimate of your competencies to adapt the results to you.

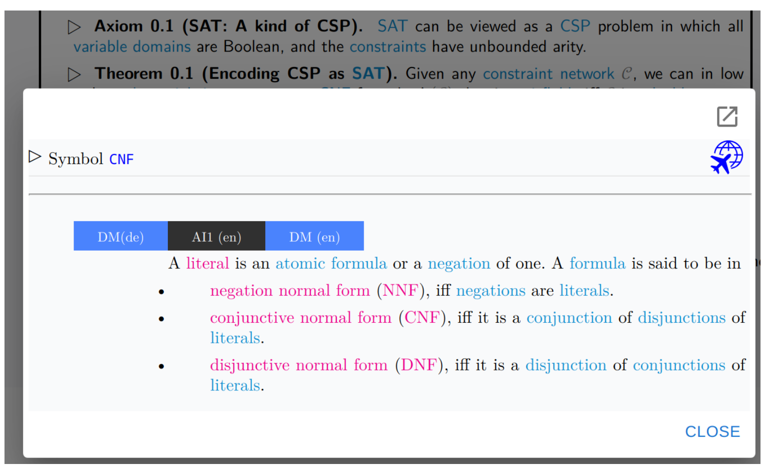
Learning Support Services in ALEA

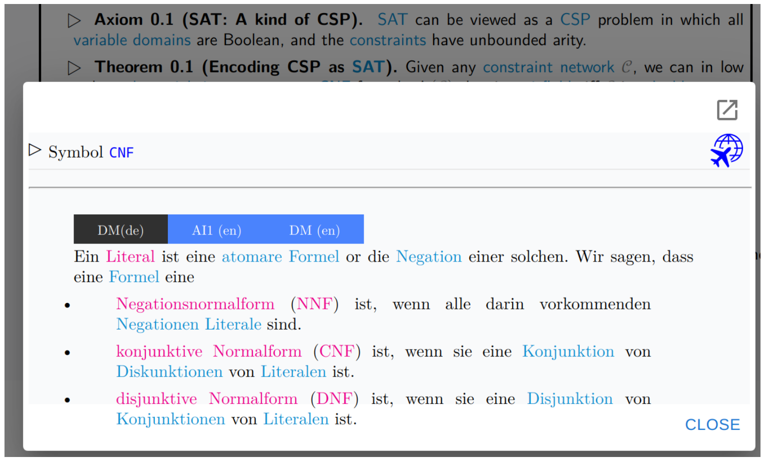
- **Idea:** Embed learning support services into active course materials.
- **Example A.1.3 (Definition on Hover).** Hovering on a (cyan) term reference reminds us of its definition. (even works recursively)



▷ **Example A.1.4 (More Definitions on Click).** Clicking on a (cyan) term reference shows us more definitions from other contexts.







▷ **Example A.1.5 (Guided Tour).** A guided tour for a concept c assembles definitions/etc. into a self-contained mini-cours

c = countable ⇝

▷ . . . your idea here . . . (the sky is the limit)

FAU : 736 2025-05-01

Note that this is only an initial collection of learning support services, we are constantly working on additional ones. Look out for feature notifications (

) on the upper right hand of the ALeA screen.

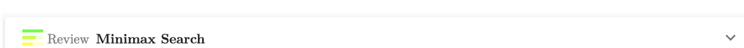
(Practice/Remedial) Problems Everywhere

▷ **Problem:** Learning requires a mix of understanding and test-driven practice.

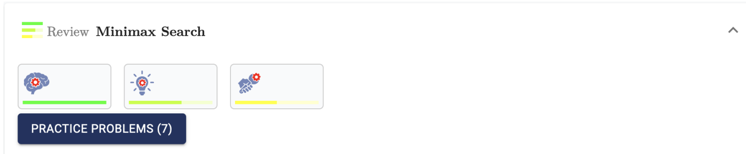
▷ **Idea:** ALeA supplies targeted practice problems everywhere.

▷ **Concretely:** Revision markers at the end of sections.

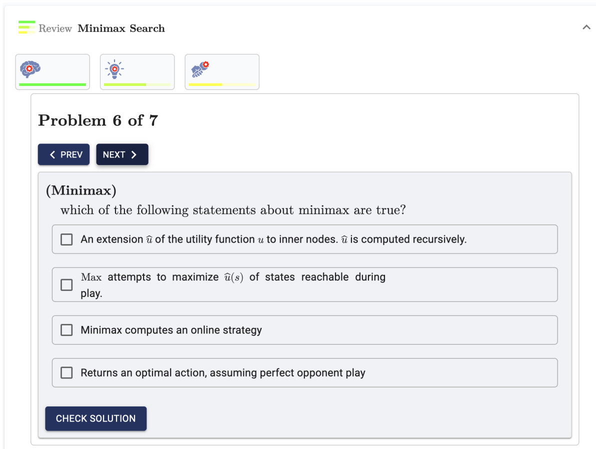
- ▷ A relatively non-intrusive overview over competency



- ▷ Click to extend it for details.



▷ Practice problems as usual. (targeted to your specific competency)



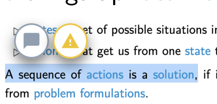
While the learning support services up to now have been adressed to individual learners, we now turn to services addressed to communities of learners, ranging from study groups with three learners, to whole courses, and even – eventually – all the alumni of a course, if they have not de-registered from ALeA.

Currently, the community aspect of ALeA only consists in localized interactions with the course materials.
The ALeA system uses the semantic structure of the course materials to localize some interactions that are otherwise often from separate applications. Here we see two:

1. one for reporting content errors – and thus making the material better for all learners – and‘’
2. a localized course forum, where forum threads can be attached to learning objects.

Localized Interactions with the Community

▷ Selecting text brings up localized – i.e. anchored on the selection – interactions:



▷ post a (public) comment or take (private) note

▷ report an error to the course authors/instructors

▷ Localized comments induce a thread in the ALeA forum (like the StudOn Forum, but targeted towards specific learning objects.)

▷ Answering questions gives karma $\widehat{=}$ a public measure of user helpfulness.

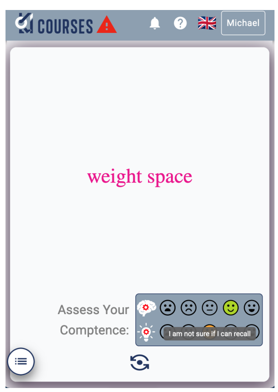
▷ Notes can be anonymous (⇝ generate no karma)

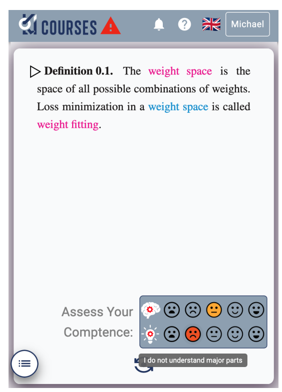
FAU : 738 2025-05-01

We can use the same four models discussed in the space of guided tours to deploy additional learning support services, which we now discuss.

New Feature: Drilling with Flashcards

▷ Flashcards challenge you with a task (term/problem) on the front...





...and the definition/answer is on the back.

▷ Self-assessment updates the learner model (before/after)

▷ **Idea:** Challenge yourself to a card stack, keep drilling/assessing flashcards until the learner model eliminates all.

▷ **Bonus:** Flashcards can be generated from existing semantic markup (educational equivalent to free beer)

FAU : 739 2025-05-01

We have already seen above how the learner model can drive the drilling with flashcards. It can also be used for the configuration of card stacks by configuring a domain e.g. a section in the course materials and a competency threshold. We now come to a very important issue that we always face when we do AI systems that interface with humans. Most web technology companies that take one the approach "the user pays for the services with their personal data, which is sold on" or integrate advertising for renumeration. Both are not acceptable in university setting.

But abstaining from monetizing personal data still leaves the problem how to protect it from intentional or accidental misuse. Even though the GDPR has quite extensive exceptions for research, the ALeA system – a research prototype – adheres to the principles and mandates of the GDPR. In particular it makes sure that personal data of the learners is only used in learning support services directly or indirectly initiated by the learners themselves.

Learner Data and Privacy in ALEA

▷ **Observation:** Learning support services in ALEA use the learner model; they
  - ▷ need the learner model data to adapt to the invidivual learner!
  - ▷ collect learner interaction data (to update the learner model)

▷ **Consequence:** You need to be logged in (via your FAU IDM credentials) for useful learning support services!

▷ **Problem:** Learner model data is highly sensitive personal data!

▷ **ALeA Promise:** The ALEA team does the utmost to keep your personal data safe. (SSO via FAU IDM/eduGAIN, ALEA trust zone)

▷ **ALeA Privacy Axioms:**

1. ALEA only collects learner models data about logged in users.
2. Personally identifiable learner model data is only accessible to its subject (delegation possible)
3. Learners can always query the learner model about its data.
4. All learner model data can be purged without negative consequences (except usability deterioration)
5. Logging into ALEA is completely optional.

▷ **Observation:** Authentication for bonus quizzes are somewhat less optional, but you can always purge the learner model later.

FAU : 740 2025-05-01

So, now that you have an overview over what the ALEA system can do for you, let us see what you have to concretely do to be able to use it.

Concrete Todos for ALeA

▷ **Recall:** You will use ALeA for the prepquizzes (or lose bonus points)
All other use is optional. (but AI-supported pre/postparation can be helpful)

▷ To use the ALeA system, you will have to log in via SSO: (do it now)
  - ▷ go to `https://courses.voll-ki.fau.de/course-home/ai-2`,
  - ▷ in the upper right hand corner you see 

,
  - ▷ log in via your FAU IDM credentials. (you should have them by now)

- ▷ You get access to your personal ALeA profile via
  (plus feature notifications, manual, and language chooser)

▷ **Problem:** Most ALeA services depend on the learner model. (to adapt to you)

▷ **Solution:** Initialize your learner model with your educational history!

- ▷ **Concretely**: enter taken CS courses (FAU equivalents) and grades.
- ▷ ALeA uses that to estimate your CS/AI competencies. (for your benefit)
- ▷ then ALeA knows about you; I don't! (ALeA trust zone)

Even if you did not understand some of the AI jargon or the underlying methods (yet), you should be good to go for using the ALeA system in your day-to-day work.