1 Introduction

1.1 What is AI?

Literature:

- Empfohlenes Begleitbuch: Russel and Norvig, Artificial Intelligence: A Modern Approach, 4. Edition 2020.

1.1.1 Definitionen (Definitions)

1.1.2 Definitions

There is no easy, official definition for AI. Two classic definitions are:

- **John McCarthy (1971):** "The science and engineering of making intelligent machines, especially intelligent computer programs." AI does not have to confine itself to methods that are biologically observable.
- **Marvin Minsky (1969):** "The science of making machines do things that would require intelligence if done by men".

1.1.3 Categories of AI

AI definitions can be classified along two dimensions

1. Thought processes/reasoning vs. behavior/action
2. Success according to human standards vs. success according to an ideal concept of intelligence (rationality)

- **Systems that think like humans:**
  - Cognitive Science.
  - Builds on cognitive models validated by psychological experiments and neurological data.
- **Systems that act like humans:**
  - The **Turing Test**
- **Systems that think rationally:**
  - Focus on "Laws of Thoughts," correct argument processes.
- **Systems that act rationally:**
  - Focus on "doing the right thing" (**Rational Behavior**).
  - A rationally acting system maximizes the achievement of its goals based on the available information.
  - This is more general than rational thinking (as a provably correct action often does not exist) and more amenable to analysis.

1.1.4 General vs. Narrow AI

- **General (Strong) AI:** Can handle *any* intellectual task that a human can. This is a research goal.
- **Narrow (Weak) AI:** Is specified to deal with a *concrete* or a set of specified tasks. This is what we currently use primarily.

1.2 What is Intelligence?

1.2.1 The Turing Test

- **Question:** When does a system behave intelligently?

- **Assumption:** An entity is intelligent if it cannot be distinguished from another intelligent entity by observing its behavior.
- **Test:** A human interrogator interacts "blind" (e.g., via text) with two players (A and B), one of whom is a human and one a computer.
- **Goal:** If the interrogator cannot determine which player... is a computer... the computer is said to pass the test.
- **Relevance:** The test is still relevant, requires major components of AI (knowledge, reasoning, language, learning), but is hard/not reproducible and not amenable to mathematical analysis.

1.2.2 The Chinese Room Argument

- **Question:** Is intelligence the same as intelligent behavior?
- **Assumption:** Even if a machine behaves in an intelligent manner, it does not have to be intelligent at all (i.e., without understanding).
- **Thought Experiment:** A person who doesn't know Chinese is locked in a room. They receive Chinese notes (questions) and have a detailed instruction book telling them which Chinese symbols (answers) to output based on the input symbols, without understanding it at all.
- **Result:** From the outside, the room "understands" Chinese (it behaves intelligently), but the person inside understands nothing.
- **Follow-up Question:** Is a self-driving car intelligent?

1.3 Foundations, Taxonomy & Limits

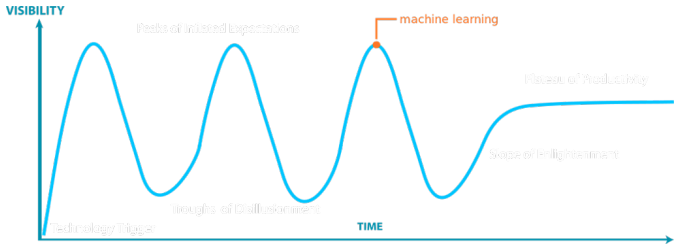
1.3.1 Foundations of AI

AI is an interdisciplinary field built on contributions from many areas:

- **Philosophy:** Logic, reasoning, rationality, mind as a physical system.
- **Mathematics:** Formal representation and proof, computation, 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:** ...optimal agent design.

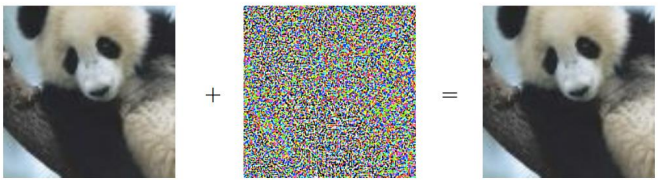
1.3.2 Taxonomy and History

- **Taxonomy: Artificial Intelligence** is the broadest field. **Machine Learning (ML)** is a subfield of AI. **Deep Learning** is a subfield of ML.
- **Subdisciplines of AI:** Include Machine Learning, Deep Learning, Search and Optimization, Robotics, Natural Language Processing (NLP), Computer Vision (CV), and Cognitive Science.
- **History:** The development of AI occurred in cycles, often called "AI Winters". Hype phases ("Peaks of Inflated Expectations") existed for "neural networks", "expert systems", and "machine learning".



1.3.3 Limits of Current AI

- **"A.I. is harder than you think":**
  - Current AI is often isolated to single problems.
  - AI models are **not without bias**.
  - There are **fundamental differences** in how AI perceives the world/environment.
- **AI can be tricked (Adversarial Examples):**
  - AI systems can be manipulated by perturbations (noise) often invisible to humans.
  - Example: An image of a "panda" is classified as a "gibbon" with high confidence after adding noise.



"panda"
57.7% confidence

"gibbon"
99.3 % confidence