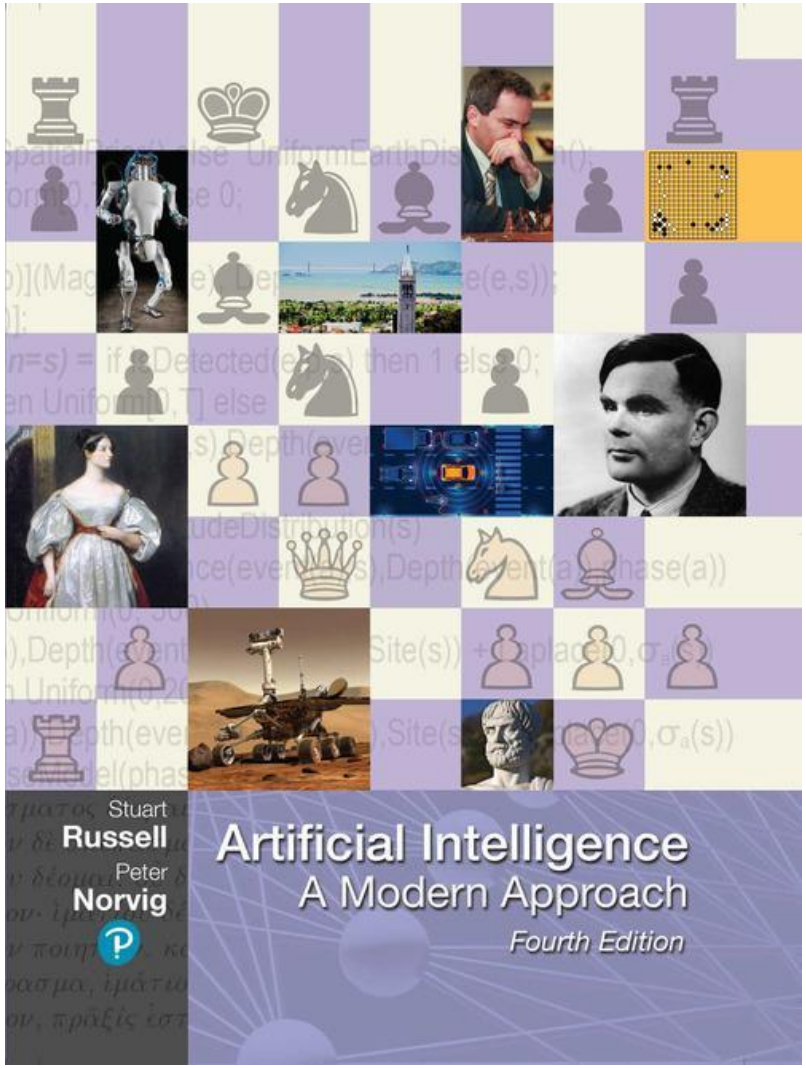
Artificial Intelligence: A Modern Approach

Fourth Edition



Chapter 28

The Future Of AI

Outline

- AI Components
- AI Architectures

AI Components

Sensors and actuators

- direct access to the world has been absent
- AI systems were built in such a way that humans had to supply the inputs and interpret the outputs.

- The demand for better image processing in cellphone cameras has given us inexpensive high-resolution cameras for use in robotics

- MEMS (micro-electromechanical systems) technology has supplied miniaturized accelerometers, gyroscopes, and actuators small enough to fit in artificial flying insects

- AI systems are at the cusp of moving from primarily software-only systems to useful embedded robotic system

- intelligent robots will first make strides in industry before the home market

AI Components

Representing the state of the world

- Keeping track of the world requires perception as well as updating of internal representations
- Current filtering and perception algorithms can be combined to do a reasonable job of recognizing objects and reporting low-level predicates
- Future progress will require techniques that generalize to novel situations without requiring exhaustive examples
- word embeddings and similar representations can free us from the strict bounds of concepts defined by necessary and sufficient condition
- However it remains a daunting task to define general, reusable representation schemes for complex domains.

AI Components

Selecting actions

- The primary difficulty in action selection in the real world is coping with long-term plans
- Humans apply hierarchical structure on behavior
- hierarchical reinforcement learning has succeeded in combining these ideas with the MDP formalism
- these methods have not been extended to the partially observable case (POMDPs).

Deciding what we want

- the task of picking the right utility function is a challenging problem
- do not yet have much experience with building complex real-world preference models, let alone probability distributions over such models.
- **Inverse reinforcement learning** is one approach when we have an expert who can perform a task, but not explain it.
- We need better ways of saying what we want and better ways for robots to interpret the information we provide.
- powerful ecosystem for aggregating user preferences but fail to provide an easy way of opting out
- in the future we will have personal agents that stick up for our true long-term interests

AI Components

Learning

- Current algorithms can cope with quite large problems, reaching or exceeding human capabilities in many task
- learning can stall when data are sparse, or unsupervised, or when we are dealing with complex representation
- need advances in transfer learning so that we can take advantage of data in one domain to improve performance on a related domain
- The vast majority of machine learning research today assumes a factored representation

AI Components

Resources

- Machine learning research and development has been accelerated by the increasing availability of data, storage, processing power, software, trained experts, and the investments needed to support them
- Hundreds of high-quality data sets are available for a range of tasks in computer vision, speech recognition, and natural language processing.
- There is a possibility that quantum computers could accelerate AI.
- Currently there are some fast quantum algorithms for the linear algebra operations used in machine learning
- Current quantum computers handle only a few tens of bits, whereas machine learning algorithms often handle inputs with millions of bits and create models with hundreds of millions of parameters.

AI Architectures

- AI has long had a split between symbolic systems (based on logical and probabilistic inference) and connectionist systems (based on loss minimization over a large number of uninterpreted parameters).
- Agents need ways to control their own deliberations
  - use the available time well
  - cease deliberating when action is demanded
- anytime algorithms
  - an algorithm whose output quality improves gradually over time, so that it has a reasonable decision ready whenever it is interrupted
- Decision-theoretic metareasoning
  - applies the theory of information value to the selection of individual computations
- Metareasoning techniques can be used to design better search algorithms and to guarantee that the algorithms have the anytime property
- **Reflective architecture**: an architecture that enables deliberation about the computational entities and actions occurring within the architecture itself.

AI Architectures

General AI

- Much of the progress in AI in the 21st century so far has been guided by competition on narrow tasks, such as the DARPA Grand Challenge for autonomous cars, the ImageNet object recognition competition

- Continued work on specific tasks (or on individual components) will not be enough to reach mastery on a wide variety of tasks

- AI as a field has made a reasonable exploration/exploitation tradeoff, assembling a portfolio of components, improving on particular tasks, while also exploring promising and sometimes far-out new ideas.

- Work on components can spur new ideas; for example, generative adversarial networks (GANs) and transformer language models each opened up new areas of research.

AI Architectures

AI engineering

- The AI industry has not yet reached that level of maturity.
- We do have a variety of powerful tools and frameworks, such as TensorFlow, Keras, PyTorch, CAFFE, Scikit-Learn and SCIPY.
- But many of the most promising approaches, such as GANs and deep reinforcement learning, have proven to be difficult to work with—they require experience and a degree of fiddling to get them to train properly in a new domain
- start with a single huge system and, for each new task, extract from it the parts that are relevant to the task

AI Architectures

The future

- AI seems to fit in with other powerful revolutionary technologies such as printing, plumbing, air travel, and telephony.
- AI is different from previous revolutionary technologies.
- Improving printing, plumbing, air travel, and telephony to their logical limits would not produce anything to threaten human supremacy in the world.
- Improving AI to its logical limit certainly could.

In conclusion, AI has made great progress in its short history, but the final sentence of Alan Turing's (1950) essay on Computing Machinery and Intelligence is still valid today:

We can see only a short distance ahead,
but we can see that much remains to be done.