EXERCISES 1

INTRODUCTION

Note that for many of the questions in this chapter, we give references where answers can be found rather than writing them out—the full answers would be far too long.

1.1 What Is AI?

Exercise 1.1.#DEFA

Define in your own words: (a) intelligence, (b) artificial intelligence, (c) agent, (d) rationality, (e) logical reasoning.

a. Dictionary definitions of **intelligence** talk about "the capacity to acquire and apply knowledge" or "the faculty of thought and reason" or "the ability to comprehend and profit from experience." These are all reasonable answers, but if we want something quantifiable we would use something like "the ability to act successfully across a wide range of objectives in complex environments."

b. We define **artificial intelligence** as the study and construction of agent programs that perform well in a given class of environments, for a given agent architecture; they *do the right thing*. An important part of that is dealing with the uncertainty of what the current state is, what the outcome of possible actions might be, and what is it that we really desire.

c. We define an **agent** as an entity that takes action in response to percepts from an environment.

d. We define **rationality** as the property of a system which does the "right thing" given what it knows. See Section 2.2 for a more complete discussion. The basic concept is *perfect* rationality; Section **??** describes the impossibility of achieving perfect rationality and proposes an alternative definition.

e. We define **logical reasoning** as the a process of deriving new sentences from old, such that the new sentences are necessarily true if the old ones are true. (Notice that does not refer to any specific syntax or formal language, but it does require a well-defined notion of truth.)

Exercise 1.1.#TURI

Read Turing's original paper on AI (Turing, 1950). In the paper, he discusses several objections to his proposed enterprise and his test for intelligence. Which objections still carry

weight? Are his refutations valid? Can you think of new objections arising from developments since he wrote the paper? In the paper, he predicts that, by the year 2000, a computer will have a 30% chance of passing a five-minute Turing Test with an unskilled interrogator. What chance do you think a computer would have today? In another 25 years?

See the solution for exercise 26.1 for some discussion of potential objections.

The probability of fooling an interrogator depends on just how unskilled the interrogator is. A few entrants in the Loebner prize competitions have fooled judges, although if you look at the transcripts, it looks like the judges were having fun rather than taking their job seriously. There certainly have been examples of a chatbot or other online agent fooling humans. For example, see the description of the Julia chatbot at `www.lazytd.com/lti/julia/`. We'd say the chance today is something like 10%, with the variation depending more on the skill of the interrogator rather than the program. In 25 years, we expect that the entertainment industry (movies, video games, commercials) will have made sufficient investments in artificial actors to create very credible impersonators.

Note that governments and international organizations are seriously considering rules that require AI systems to be identified as such. In California, it is already illegal for machines to impersonate humans in certain circumstances.

Exercise 1.1.#REFL

Are reflex actions (such as flinching from a hot stove) rational? Are they intelligent?

Yes, they are rational, because slower, deliberative actions would tend to result in more damage to the hand. If "intelligent" means "applying knowledge" or "using thought and reasoning" then it does not require intelligence to make a reflex action.

Exercise 1.1.#SYAI

To what extent are the following computer systems instances of artificial intelligence:

- Supermarket bar code scanners.
- Web search engines.
- Voice-activated telephone menus.
- Spelling and grammar correction features in word processing programs.
- Internet routing algorithms that respond dynamically to the state of the network.

- Although bar code scanning is in a sense computer vision, these are not AI systems. The problem of reading a bar code is an extremely limited and artificial form of visual interpretation, and it has been carefully designed to be as simple as possible, given the hardware.
- In many respects. The problem of determining the relevance of a web page to a query is a problem in natural language understanding, and the techniques are related to those

we will discuss in Chapters 23 and 24. Search engines also use clustering techniques analogous to those we discuss in Chapter 20. Likewise, other functionalities provided by a search engines use intelligent techniques; for instance, the spelling corrector uses a form of data mining based on observing users' corrections of their own spelling errors. On the other hand, the problem of indexing billions of web pages in a way that allows retrieval in seconds is a problem in database design, not in artificial intelligence.

- To a limited extent. Such menus tends to use vocabularies which are very limited – e.g. the digits, "Yes", and "No" — and within the designers' control, which greatly simplifies the problem. On the other hand, the programs must deal with an uncontrolled space of all kinds of voices and accents. Modern digital assistants like Siri and the Google Assistant make more use of artificial intelligence techniques, but still have a limited repetoire.
- Slightly at most. The spelling correction feature here is done by string comparison to a fixed dictionary. The grammar correction is more sophisticated as it need to use a set of rather complex rules reflecting the structure of natural language, but still this is a very limited and fixed task.

  The spelling correctors in search engines would be considered much more nearly instances of AI than the Word spelling corrector are, first, because the task is much more dynamic – search engine spelling correctors deal very effectively with proper names, which are detected dynamically from user queries – and, second, because of the technique used – data mining from user queries vs. string matching.
- This is borderline. There is something to be said for viewing these as intelligent agents working in cyberspace. The task is sophisticated, the information available is partial, the techniques are heuristic (not guaranteed optimal), and the state of the world is dynamic. All of these are characteristic of intelligent activities. On the other hand, the task is very far from those normally carried out in human cognition. In recent years there have been suggestions to base more core algorithmic work on machine learning.

Exercise 1.1.#COGN

Many of the computational models of cognitive activities that have been proposed involve quite complex mathematical operations, such as convolving an image with a Gaussian or finding a minimum of the entropy function. Most humans (and certainly all animals) never learn this kind of mathematics at all, almost no one learns it before college, and almost no one can compute the convolution of a function with a Gaussian in their head. What sense does it make to say that the "vision system" is doing this kind of mathematics, whereas the actual person has no idea how to do it?

Presumably the brain has evolved so as to carry out this operations on visual images, but the mechanism is only accessible for one particular purpose in this particular cognitive task of image processing. Until about two centuries ago there was no advantage in people (or animals) being able to compute the convolution of a Gaussian for any other purpose.

The really interesting question here is what we mean by saying that the "actual person" can do something. The person can see, but he cannot compute the convolution of a Gaussian;

but computing that convolution is *part* of seeing. This is beyond the scope of this solution manual.

Exercise 1.1.#EVOR

Why would evolution tend to result in systems that act rationally? What goals are such systems designed to achieve?

The notion of acting rationally *presupposes* an objective, whether explicit or implicit. We understand evolution as a process that operates in the physical world, where there are no inherent objectives. So the question is really asking whether evolution tends to produce systems whose behavior can be interpreted consistently as rational according to some objective.

It is tempting to say that evolution tends to produce organisms that act rationally in the pursuit of reproduction. This is not completely wrong but the true picture is much more complex because of the question of what "system" refers to—it could be organisms (humans, rats, bacteria), superorganisms (ant and termite colonies, human tribes, corals), and even individual genes and groups of genes within the genome. Selection and mutation processes operate at all these levels. By definition, the systems that exist are those whose progenitors have reproduced successfully. If we consider an ant colony, for example, there are many individual organisms (e.g., worker ants) that do not reproduce at all, so it is not completely accurate to say that evolution produces organisms whose objective is to reproduce.

Exercise 1.1.#AISC

Is AI a science, or is it engineering? Or neither or both? Explain.

This question is intended to be about the essential nature of the AI problem and what is required to solve it, but could also be interpreted as a sociological question about the current practice of AI research.

A *science* is a field of study that leads to the acquisition of empirical knowledge by the scientific method, which involves falsifiable hypotheses about what is. A pure *engineering* field can be thought of as taking a fixed base of empirical knowledge and using it to solve problems of interest to society. Of course, engineers do bits of science—e.g., they measure the properties of building materials—and scientists do bits of engineering to create new devices and so on.

The "human" side of AI is clearly an empirical science—called cognitive science these days—because it involves psychological experiments designed out to find out how human cognition actually works. What about the the "rational" side? If we view it as studying the abstract relationship among an arbitrary task environment, a computing device, and the program for that computing device that yields the best performance in the task environment, then the rational side of AI is really mathematics and engineering; it does not require any empirical knowledge about the *actual* world—and the *actual* task environment—that we inhabit; that a given program will do well in a given environment is a *theorem*. (The same is true of pure decision theory.) In practice, however, we are interested in task environments that do approximate the actual world, so even the rational side of AI involves finding out what the actual

world is like. For example, in studying rational agents that communicate, we are interested in task environments that contain humans, so we have to find out what human language is like. In studying perception, we tend to focus on sensors such as cameras that extract useful information from the actual world. (In a world without light, cameras wouldn't be much use.) Moreover, to design vision algorithms that are good at extracting information from camera images, we need to understand the actual world that generates those images. Obtaining the required understanding of scene characteristics, object types, surface markings, and so on is a quite different kind of science from ordinary physics, chemistry, biology, and so on, but it is still science.

In summary, AI is definitely engineering but it would not be especially useful to us if it were not also an empirical science concerned with those aspects of the real world that affect the design of intelligent systems for that world.

Exercise 1.1.#INTA

"Surely computers cannot be intelligent—they can do only what their programmers tell them." Is the latter statement true, and does it imply the former?

This depends on your definition of "intelligent" and "tell." In one sense computers only do what the programmers command them to do, but in another sense what the programmers consciously tells the computer to do often has very little to do with what the computer actually does. Anyone who has written a program with an ornery bug knows this, as does anyone who has written a successful machine learning program. So in one sense Samuel "told" the computer "learn to play checkers better than I do, and then play that way," but in another sense he told the computer "follow this learning algorithm" and it learned to play. So we're left in the situation where you may or may not consider learning to play checkers to be a sign of intelligence (or you may think that learning to play in the right way requires intelligence, but not in this way), and you may think the intelligence resides in the programmer or in the computer.

Exercise 1.1.#INTB

"Surely animals cannot be intelligent—they can do only what their genes tell them." Is the latter statement true, and does it imply the former?

The point of this exercise is to notice the parallel with the previous one. Whatever you decided about whether computers could be intelligent in 1.11, you are committed to making the same conclusion about animals (including humans), *unless* your reasons for deciding whether something is intelligent take into account the mechanism (programming via genes versus programming via a human programmer). Note that Searle makes this appeal to mechanism in his Chinese Room argument (see Chapter 27).

Exercise 1.1.#INTC

"Surely animals, humans, and computers cannot be intelligent—they can do only what their constituent atoms are told to do by the laws of physics." Is the latter statement true, and does it imply the former?

Again, your definition of "intelligent" drives your answer to this question.

1.2 The Foundations of Artificial Intelligence

Exercise 1.2.#NTRC

There are well-known classes of problems that are intractably difficult for computers, and other classes that are provably undecidable. Does this mean that AI is impossible?

No. It means that AI systems should avoid trying to solve intractable problems. Usually, this means they can only approximate optimal behavior. Notice that humans don't solve NP-complete problems either. Sometimes they are good at solving specific instances with a lot of structure, perhaps with the aid of background knowledge. AI systems should attempt to do the same.

Exercise 1.2.#SLUG

The neural structure of the sea slug *Aplysia* has been widely studied (first by Nobel Laureate Eric Kandel) because it has only about 20,000 neurons, most of them large and easily manipulated. Assuming that the cycle time for an *Aplysia* neuron is roughly the same as for a human neuron, how does the computational power, in terms of memory updates per second, compare with the personal computer described in Figure 1.2?

Depending on what you want to count, the computer has a thousand to a million times more storage, and a thousand times more operations per second.

Exercise 1.2.#INTR

How could introspection—reporting on one's inner thoughts—be inaccurate? Could I be wrong about what I'm thinking? Discuss.

Just as you are unaware of all the steps that go into making your heart beat, you are also unaware of most of what happens in your thoughts. You do have a conscious awareness of some of your thought processes, but the majority remains opaque to your consciousness. The field of psychoanalysis is based on the idea that one needs trained professional help to analyze one's own thoughts. Neuroscience has also shown that we are unaware of much of the activity in our brains.

1.3 The History of Artificial Intelligence

Exercise 1.3.#IQEV

Suppose we extend Evans's ANALOGY program (Evans, 1968) so that it can score 200 on a standard IQ test. Would we then have a program more intelligent than a human? Explain.

No. IQ test scores correlate well with certain other measures, such as success in college, ability to make good decisions in complex, real-world situations, ability to learn new skills and subjects quickly, and so on, but *only* if they're measuring fairly normal humans. The IQ test doesn't measure everything. A program that is specialized only for IQ tests (and specialized further only for the analogy part) would very likely perform poorly on other measures of intelligence. Consider the following analogy: if a human runs the 100m in 10 seconds, we might describe him or her as *very athletic* and expect competent performance in other areas such as walking, jumping, hurdling, and perhaps throwing balls; but we would not describe a Boeing 747 as *very athletic* because it can cover 100m in 0.4 seconds, nor would we expect it to be good at hurdling and throwing balls.

Even for humans, IQ tests are controversial because of their theoretical presuppositions about innate ability (distinct from training effects) and the generalizability of results. See *The Mismeasure of Man* (Stephen Jay Gould, 1981) or *Multiple Intelligences: the Theory in Practice* (Howard Gardner, 1993) for more on IQ tests, what they measure, and what other aspects there are to "intelligence."

Exercise 1.3.#PRMO

Some authors have claimed that perception and motor skills are the most important part of intelligence, and that "higher level" capacities are necessarily parasitic—simple add-ons to these underlying facilities. Certainly, most of evolution and a large part of the brain have been devoted to perception and motor skills, whereas AI has found tasks such as game playing and logical inference to be easier, in many ways, than perceiving and acting in the real world. Do you think that AI's traditional focus on higher-level cognitive abilities is misplaced?

Certainly perception and motor skills are important, and it is a good thing that the fields of vision and robotics exist (whether or not you want to consider them part of "core" AI). But given a percept, an agent still has the task of "deciding" (either by deliberation or by reaction) which action to take. This is just as true in the real world as in artificial micro-worlds such as chess-playing. So computing the appropriate action will remain a crucial part of AI, regardless of the perceptual and motor system to which the agent program is "attached." On the other hand, it is true that a concentration on micro-worlds has led AI away from the really interesting environments such as those encountered by self-driving cars.

Exercise 1.3.#WINT

Several "AI winters," or rapid collapses in levels of economic and academic activity (and media interest) associated with AI, have occurred. Describe the causes of each collapse and of the boom in interest that preceded it.

In addition to the information in the chapter, ? (?, ?), ? (?), and ? (?) provide ample starting material for the aspiring historian of AI. One can identify at least three AI winters (although the phrase was not applied to the first one, because the original phrase **nuclear winter** did not emerge until the early 1980s).

a. As noted in the chapter, research funding dried up in the early 1970s in both the US and UK. The ostensible reason was failure to make progress on the rather lavish promises of the 1960s, particularly in the areas of neural networks and machine translation. In 1970, the US Congress curtailed most AI funding from ARPA, and in 1973 the Lighthill report in the UK ended funding for all but a few researchers. Lighthill referred particularly to the difficulties of overcoming the combinatorial explosion.

b. In the late 1980s, the expert systems boom ended, due largely to the difficulty and expense of building and maintaining expert systems for complex applications, the lack of a valid uncertainty calculus in these systems, and the lack of interoperability between AI software and hardware and existing data and computation infrastructure in industry.

c. In the early 2000s, the end of the dot-com boom also ended an upsurge of interest in the use of AI systems in the burgeoning online ecosystem. AI systems had been used for such tasks as information extraction from web pages to support shopping engines and price comparisons; various kinds of search engines; planning algorithms for achieving complex goals requiring several steps and combining information from multiple web pages; and converting human-readable web pages into machine-readable database tuples to allow global information aggregation, as in citation databases constructed from online pdf files.

It is also interesting to explore the extent to which the winters were due to over-optimistic and exaggerated claims by AI researchers or to over-enthusiasm and over-interpretation of the significance of early results by funders and investors.

Exercise 1.3.#DLAI

The resurgence of interest in AI in the 2010s is often attributed to deep learning. Explain what deep learning is, how it relates to AI as a whole, and where the core technical ideas actually originated.

Deep learning is covered in Section 1.3.8, where it is defined as "machine learning using multiple layers of simple, adjustable computing elements." Thus, it is a particular branch of machine learning, which is itself a subfield of AI. Since many AI systems do not use learning at all, and there are many effective machine learning techniques that are unrelated to deep learning, the view (often expressed in popular articles on AI) that deep learning has "replaced" AI is wrong for multiple reasons.