0 Introduction

0.1 The Elusive Definition of (Artificial) Intelligence

The fundamental problem with defining **artificial intelligence** is that we do not have a definition of **intelligence**. We have a lot of bad attempts, mostly by psychologists and cognitive scientists. For example, here was a major public attempt:[1]

> A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings — “catching on,” “making sense” of things, or “figuring out” what to do.

This is a typical definition, and like most typical definitions, it is awful. It is so general and ill-defined that we might claim that, for example, computers are *already* intelligent, because they can *already* do these things to some degree. The issue is degree: the definition, like most others, offers no measurement or point beyond which we can claim something is intelligent. It also doesn’t define intelligence so much as suggest outward symptoms of “intelligence”. But the most fundamental problem with this definition is that it is filled with circular, undefined terms. Intelligence involves “thinking”, “comprehending”, “figuring out what to do”, “making sense of things”, and “catching on”. In other words, intelligence consists of being intelligent.

Why is this definition so vague and self-referential? Because in fact we don’t have a definition of intelligence at all. We only have a **model of intelligence**: ourselves.[2] Intelligence is basically defined in terms of stuff which humans can do but other things (living or otherwise) cannot do.

Figure 0 Clever Hans

This notion is very old indeed. What can we do that no other creature can do? What separates us from the animals? What makes us godlike? The Stoics[3] thought it was our ability to **reason**.[4] This argument has held in philosophy even to this day: dogs can’t do calculus. Cows can’t play chess. Rocks can’t prove things. Thus they are not as intelligent as we are, whatever that means. It’s what made **Clever Hans** so revolutionary in 1900: he was a horse who seemed to be able to arithmetic.[5]

[1] Arvey, Richard (and 51 other signatories), “Mainstream Science on Intelligence”, *Wall Street Journal*, December 13, 1994. They were responding to claims about intelligence being made in the book *The Bell Curve*. For more information, see http://en.wikipedia.org/wiki/Mainstream_Science_on_Intelligence

[2] And, to the degree that they resemble us and how we act, we may deign to allow certain other animals into the “intelligent” club.

[3] Greece, around 200 BC. Followers of Zeno.

[4] A different opinion: “Weaseling out of things is important to learn. It’s what separates us from the animals. Except the weasel.” (Homer Simpson, from “Boy-Scoutz n the Hood”, *The Simpsons*.)

[5] Sadly, it was not true. But it wasn’t actually fraud. Instead, it turned out he was following his master’s unintentional physical cues. Clever Hans indeed!

0.2 The Turing Test

It was into this fray the Alan Turing stepped when he proposed a wholesale rejection of a **definition** of intelligence, and to replace it with a **comparison** to our only model (namely us).[6] Turing suggested replacing the question, "Can computers think?" with an actual test which skirted the pesky issue of what "thinking" ("being intelligent) meant. Here's how it worked.

Turing proposed a game called the **imitation game**, with three players, A, B, and C. A is a man, and B is a woman. C is the *interrogator*, and C only communicates with A and B via typing remote messages. C does not know which is the man and which is the woman. It is C's job to figure out which is which. It is both A's and B's job to convince C that each is the woman. C can ask A and B anything within reason, but not things like "come meet me at 3:00 PM on the corner of 3rd and Main." If A were clever enough, and convinced C he was the woman, he'd win the game.

Now Turing asked the following. What if we replaced A with a computer? Now A's job (and B's) is to convince the interrogator that each is the *human*. What if C couldn't tell which was the real woman (or human)? What would that imply?

So let's say that we believe that humans are, generally speaking, intelligent. How do we know that Bob next door is intelligent? I suppose it's because there is some *quality* about Bob, which we can't quite put our finger on, which we ascribe to "intelligence" whatever that means. He just seems smart. Is it because of his eyebrows? His smart-looking clothing? Or because we can hear him talk? Or because he can hear us or see us? If so, what about Helen Keller?[7] Surely she was intelligent? Turing supposed that most of us would assume, reasonably, that intelligence was not a feature of what we looked or sounded like, or our ability to read or hear. So he suggested we strip all that away. His suggestion was to set up the test with remotely typed messages back and forth so as to eliminate those physical qualities. Surely we should still be able to ascertain intelligence based on typed communication with the remote subject.[8]

So Turing was arguing this:

1. If we claim there is an actual quality called *intelligence* (or as he called it, "thinking"), and that humans generally have this quality, and

2. If we claim that intelligence, can be discerned without seeing, smelling, poking, or otherwise physically interacting with a human, but

3. We can't tell the difference between a human and a computer without such physical interaction, then...

4. We must admit that the computer is intelligent.

Nifty. A way of testing intelligence without defining it. The Turing Test has endured lots of philosophical criticism, misinterpretation, and ridiculous mistaken contests involving chatbots. But my interest isn't really in whether the Turing Test was right or wrong.[9] It's in the fact that as early as 1950 mathematicians were owning up to the fact that for over two thousand years we'd made up a concept of intelligence without ever really defining it.

[6] Alan Turing. 1950. "Computing Machinery and Intelligence". *Mind* LIX (236). 433–460.

[7] Helen Keller was deaf, blind, and dumb, yet achieved a Bachelor of Arts degree and international stature. http://en.wikipedia.org/wiki/Helen_Keller

[8] Maybe the ability to communicate in language itself might not be a necessary feature of intelligence. But Turing's test was a good start.

[9] In truth, I think he was on to something.

0.3 Reason is Not (All of) Intelligence

AI is, and always has been, a grab-bag of techniques, loosely associated by their common goal of trying to get computers to do those things which separate us from the animals. Since **reason** had long been argued as *the thing* which defined this separation, much of early Artificial Intelligence centered around reasoning. Can we get computers to play chess? Can we get computers to figure out logical proofs? Can we get computers to solve difficult math problems?

Indeed reasoning was a strong focus of the first AI conference, in 1956 at Dartmouth College. One of the stars of the show was an actual computer program, *Logic Theorist*[10], which was capable of proving theorems in logic. Logic Theorist later re-proved theorems in Bertrand Russell and Alfred Whitehead's *Principia Mathematica*, including proving one more elegantly than the original authors had produced. Russell was excited about the result. But a paper about the proof, actually coauthored by Logic Theorist, was rejected by the *Journal of Symbolic Logic*.

What AI researchers have since discovered is that... computers are *really good at reasoning:*

- Programs like *Mathematica* are easily anyone's match in doing advanced symbolic integration, expression simplification, and other complex mathematical tasks.
- Deep Blue has defeated the world champion in Chess.[11] Most other games have fallen to computers as well.[12]
- DART, the Dynamic Analysis and Replanning Tool, was a logistics and decision support system which gained notoriety for its ability to produce significantly better military plans than humans could during the preparation for Operation Desert Storm. DART was so successful that in a presentation at Stanford, Victor Reis (the then DARPA director) famously stated that DART alone had paid for the *entire history* of DARPA investment in AI.

But strangely, there are things which computers can't do that practically *any human* can do. For example, as of my typing these words, there still does not exist a computer in the world which, when shown a picture of a child's room, can reliably point out the Teddy Bear.[13] But we have 1-year-old children which can do this. Clearly what we thought was the *hard part* of intelligence turned out not to be.

---

[10] Written in 1956 by Allen Newell, Herbert Simon, and J. C. Shaw. Simon went on to win a Nobel Prize in Economics.

[11] Well, sort of. It's since come out that IBM was tweaking Deep Blue between matches to adjust to Kasparov's strategies. In other words Deep Blue didn't defeat Kasparov: Deep Blue augmented by a large team of computer scientists and chess grandmasters defeated Kasparov. Put less gently: it seems to me that they cheated. Deep Blue also had unfair advantages over Kasparov. For example, the Deep Blue team had access to all of Kasparov's previous games, but he did not have access to any details about Deep Blue or examples of its playing.

[12] Probably the most famous example of this was *Chinook*, a checkers-playing program, which you can find at http://www.cs.ualberta.ca/∼chinook/ The rest of this footnote is taken wholesale from *Essentials of Metaheuristics*: Chinook was also the first program to win a the world championship in a nontrivial game. Marion Tinsley (*the* greatest human checkers player ever) wanted to play Chinook for the championship after Chinook started winning competitions. But the American and English checkers associations refused. So Tinsley forced their hand by resigning his title. They gave in, he got to play Chinook, and he won 4 to 2 with 33 ties. On the rematch four years later, after 6 ties, Tinsley withdrew because of stomach pains; and died soon thereafter of pancreatic cancer. So Chinook won, but sadly it did so by default. It's since improved to the point that Chinook likely cannot be beaten by any man or machine. But who knows if Tinsley would have won?

[13] Or even if there was such a computer: now let's place a Teddy Bear in a kitchen, then ask the computer "where's the refrigerator?" Its response will be to point out the Teddy Bear.

In hindsight, this should have been obvious: a large portion of our brains are machinery involved in pattern recognition. After all, human evolution from single-cell organisms spanned millions of years, and a great deal of that time was spent developing the capability to recognize things: being able to distinguish between plants and tigers is a crucial evolutionary skill. A much smaller amount of time was spent developing our capability to reason.

So anyway: it seems likely that reason isn't sufficient for intelligence. Nailing down machines which were truly "artificially intelligent"... has proven nontrivial.

0.4 A Definition of Artificial Intelligence as a Field

So lacking a definition of intelligence, let's weasel out of defining Artificial Intelligence as a *concept*. We can still make headway in defining it as a *field*.

There two classic notions of AI as a research pursuit, known as **Strong AI** and **Weak AI**. Here are my definitions of them.

> **Weak AI** Artificial Intelligence is the study of algorithms which enable computers to do tasks which previously only we humans (or higher-order animals), generally speaking, could perform because we possess Big Brains.

> **Strong AI** Artificial Intelligence is the pursuit of research leading to the development of facsimiles of the human mind.

Hollywood *loves* Strong AI. From it we get characters like Lt. Commander Data, the Terminator, and so on. It's tantalizing. But the Strong AI research community is in fact microscopically small, and consists almost entirely of armchair philosophers. There are real, serious people in the Strong AI pusuit, but there are also an amazing number of quacks.

Hollywood doesn't love Weak AI so much. It's cool but not as eye-popping as Strong AI. But Weak AI is *real* AI. It is now a huge field, encompassing everything from data mining to satellite imagery recognition to automatic language translation to GPS trip routing.[14]

Notice that my definition of Weak AI is intentionally vague. I did *not* say that computers have to do the tasks *in the same way* as we do, or perform the same *mental machinations that we humans do* when we go about these tasks. I did not say that these tasks couldn't be solved in some way other than how we Use Our Big Brains to do them. I did not say that by figuring out how to do these tasks we'd gain any insight into *how* we do it. And I didn't say that the computers in any way would necessarily *think*, whatever that means.

AI and Rationality Stewart Russell and Peter Norvig[15] like to further view the pursuit of AI as not only Weak AI but one of *rational agents*, by which they mean computational entities which act in an optimal or effectively optimal fashion in their environments. This is essentially an argument from the viewpoint of optimization. They make a strong argument for their case, but I have some qualms. First, it is very broad: by this definition we may say that trivial things like gradient descent are AI: I think AI is categorized by at least some degree of cleverness. Second, it suggests that solutions which are *satisficing* rather than optimizing are not sufficient to be called AI.

[14] This is not to say we Weak AI researchers don't secretly wish we were achieving Strong AI! We'd also like to fly spaceships to the Andromeda galaxy too.

[15] Stewart Russell and Peter Norvig. 2009. *Artificial Intelligence: a Modern Approach*. Prentice Hall.

Things that are Not AI One thing that *is not* Artificial Intelligence is so-called **Computer Game AI**. Every computer game seems to have the word "AI" emblazoned on all its advertising and packaging materials. By "AI", games usually mean simple hard-coded agent behaviors, for example hierarchical finite-state automata, designed to give the *appearance* of intelligent behavior within the context of the game.[16] Along these lines, **chatbots** have a long tradition of a quasi-natural-language-processing designed not to perform communication tasks per se but to fool people with a thin veneer of anthropomorphization. This stuff is interesting in its own right, but it's not AI.

0.5 Major Areas in Artificial Intelligence

Quite a number of techniques and research areas may reasonably fall under AI. And in addition to its own home-grown topics, AI steals liberally from probability and statistics,[17] game theory, biology, the cognitive and social sciences, and logic. There are many of ways of dividing up this seemingly arbitrary blob of topics. Here's how I do it: much of Artificial Intelligence can fall into roughly four pursuits: **Optimization**, **Induction**, **Deduction**, and **Interaction**.

Optimization is the process of wandering through an environment, hunting for samples which have the highest quality. It is *not* necessarily assumed that you will find the *best* possible item: you'll be satisfied with the best one you can find. For example, you may be looking for the price point of a product which gives the highest profit according to some economic simulation you've built. Or you could be searching for a team robot soccer strategy which performs as well as possible.

There is of course a whole subfield of mathematics involved in optimizing well-formed function: but perhaps the optimization area most closely associated with artificial intelligence is **stochastic optimization**, of which well-known examples include **metaheuristics** techniques such as the genetic algorithm, simulated annealing, or ant colony optimization.

Another optimization area which is closely associated with AI is **adaptation**: developing a program which stays on top of things. When the environment changes, it tries to adjust itself to optimize for the new environment as well as can be hoped for.

Induction is the process of producing **hypotheses** from samples observed in the environment.[18] A hypothesis, sometimes known as a **model**, is something which explains the samples that have been gathered, and ideally explains them well enough that it can **predict** the likely values of future samples. A hypothesis which is good at predicting any sample you throw at it, based on a relatively small number of initial samples it was fed, is said to be **general**. The area of AI involved in induction is known as **machine learning**. A closely related area is that of **data mining**.

---

[16] Similarly, the hoardy $A\star$ algorithm has long been co-opted for a special limited function in doing path-planning in games. This is a weak hook on which to hang a claim of using AI. This is not to say that there aren't major games out there with honest-to-goodness advanced artificial intelligence algorithms in use. But it's not all that common.

[17] There are a lot of interconnections here. Indeed, many of the areas in AI have very close analogues, at an abstract level, in statistics. AI also has very close ties to another area known as **operations research**.

[18] Not to be confused with *mathematical induction*, a proof method used for claims over infinite sets of numbers. Often this is by first by demonstrating a *base case* for some value b, then an *inductive step* which shows that if the claim holds for some $n > b$, then it must also hold for $n + 1$. For example, we may use mathematical induction to prove that all positive numbers ≥ 3 are less than the square of their immediately previous number. Our base case is simple: $3 < 2^2 = 4$. Now we show that this claim holds for every value $n > 3$. We assume that $n < (n-1)^2 = n^2 - 2n + 1$. Thus $n + 2n - 1 < n^2$. Since $n > 3$, therefore $2n - 1 > 5 > 1$, and so we may also say $n + 1 < n^2$.

Anyway, none of this has anything to do with induction in AI.

Deduction is the process of deriving new conclusions through the application of an existing set of hypotheses and rules. This is the classic process most closely associated with **reasoning**.[19] The most famous area in AI involved in deduction is **search**, the process of hunting for solutions to problems.[20] The most famous example is **state-space search**, a technique used for path-planning, proving mathematical claims, and so on. Related to state-space search is **constraint satisfaction search**. The biggest difference between the two is in the end product. State-space search often isn't interested in the solution so much in how you got there, that is, the **proof**. Whereas constraint satisfaction search is often primarily interested in the solution itself. A third common search area is **adversarial** or **game-playing search**, where the objective is to search for good plays in games or other interactive scenarios.[21]

Much of state-space search is intricately interconnected with **logic** and with various forms of **logical reasoning**. This has given rise to, essentially, advanced database systems called **knowledge representation systems** which perform sophisticated queries using logic or other reasoning methods. Though it is vitally important in artificial intelligence history, lately logic has taken a back seat to **probabilistic reasoning**, the use of probabilistic models such as Bayes Networks to produce answers to queries.

Interaction is the catch-all term I use to describe techniques used to interact with the world, with people, and with other artificial intelligence programs. This includes **computer vision**, where the system attempts to make sense, or at least make use, of what it is seeing. Vision isn't the only relevant sensory mechanism: for example **speech recognition** tries to extract linguistic meaning from raw audio. This leads to **natural language processing**, which tries to parse and understand human languages in written or spoken form.

A great deal of the computer science portion of **robotics**, and particularly **autonomous robotics** falls directly under the aegis of AI: it entails ways by which robots may autonomously (that is, on their own) interact with the world to perform the tasks they're called upon to do.

Last but not least, there is a burgeoning area of AI involving more than one artificial intelligence entity interacting with one another. This area, known as **multiagent systems**, often assumes that the entities are operating under constraints which prevent them from fully communicating or gaining a complete and global understanding of the scenario: as a result, they tend to step on each others' toes a lot, making coordination a challenge.

These are hardly hard-and-fast categories. For example, machine learning makes heavy use of optimization. Probabilistic Reasoning is closely associated with machine learning techniques for building probabilistic models. Multiagent systems which learn about one another or collectively learn about their environment (that is, perform induction) are said to be performing **multiagent learning**. Robotics makes extensive use of optimization, induction, and deduction, not to mention sibling areas in interaction. And so on. And there are "friends of the AI family" which don't fit under any of these umbrellas, such as **artificial life**. So don't take these areas as gospel.

---

[19] Though Sherlock Holmes claimed to be doing deduction, in fact, he was mostly doing induction.

[20] This is distinguished from optimization in that search is all-or-nothing: either you found the answer or you didn't. Optimization instead hunts for *as good an answer as you can get*. It's a small but important distinction which results in very diverging applications.

[21] Though it's always lumped in with state-space search, this isn't a really search method. It's an optimization method.