Foreword

This book is about the creation of super-intelligent thinking machines. The first section presents the overall case that intelligent thinking machines are not only possible but inevitable.

Then I present a model of capabilities that a system needs in order to appear intelligent, and the behaviors we can expect from a system built following that model. The details of the explanation are a bit more technical but I have endeavored to include examples which will make the process clear.

The final section extrapolates the behaviors that result from a system created along the lines of the model of Section II so we can reach conclusions about what such machines will be like and what we might do to coexist with them. It isn't critical to the thesis of this book that the model be correct in every detail. In fact, any goal-oriented learning system which interacts with our physical environment is likely to exhibit similar behavior.

What is the point of this book?

- To show that computers more intelligent than humans are possible.
- To explain why such computers are inevitable.
- To argue that machine intelligence will be created sooner than most people think.
- To demonstrate that, subsequently, vastly more powerful intelligences will be created only a few decades later.
- To conclude that such "genius" machines will lead to options and opportunities for how humans will coexist with (and prepare for) them.

As you continue through this book, you'll see a block diagram of intelligence in terms of capabilities which you can observe for yourself. The conclusion is that a reasonably sized software project can implement everything which we know about human intelligence—a fact which I'll reinforce later. Underlying this book is my contention that human intelligence is not as complex as it appears. Rather, it is built of a few fundamental capabilities, operating on an immense scale within your brain.

Over the next chapters, I intend to prove it to you. Not only that, but I make some predictions on how future intelligent machines will behave—how they will be similar to human intelligence and how they will necessarily be different. Based on these predictions, we will be able to consider how such computers and people will coexist.

AI today

Recent developments in AI (Artificial Intelligence) have been astonishing. In 1997, IBM's Deep Blue supercomputer system beat the World Chess Champion Garry Kasparov. In 2014, IBM's Watson beat champions at the TV game, *Jeopardy!* In 2015, Alphabet/Google's AlphaGo program began beating world-class players at the ancient Chinese game of Go. What's more astounding is that the October 2017 version, AlphaGo Zero, was not programmed to play Go. It was programmed to *learn* to play. And over a period of just three days of learning, playing against itself, it was able to achieve such a level of play that it could consistently beat the 2015 version.

Other fields of AI research, including speech recognition, computer vision, robotics, self-driving cars, data mining, neural networks, and deep learning, have had equally impressive successes. But are such systems intelligent? The general consensus is that they are not (although this is a matter of how we define "intelligent"). When applied to a problem outside their specific field of "expertise", most systems fail miserably. Many people use the evidence that AI has not achieved the holy grail of true general intelligence over the past 70 years as proof that either (a) true intelligence in machines is impossible or (b) true intelligence in machines is still a long way off. I disagree with both contentions.

Because of the generally limited scope of AI applications, the AI community has adopted the term AGI (Artificial General Intelligence, also called "strong AI" or "full AI"). This represents the idea of a true "thinking" machine and might represent an agglomeration of many AI technologies of more limited domain or entirely new technologies.

Why not yet? AI to AGI

Why hasn't AI already morphed into AGI? There are three primary reasons:

1. Computers have not been powerful enough to solve the problems.
2. The problems to be solved in creating intelligent systems turned out to be a lot more difficult than they initially appeared.
3. We do not yet know fully how human intelligence works.

In the next few chapters, I'll show why these roadblocks will be going away soon. I'll also expand on these and a host of other issues which have confronted the AI community.

Bringing it all together

In summary, AI has lots of bits of intelligence, but none has any underlying "understanding". I contend that AI programs have (mostly) been developed to solve specific problems. They have no contact with the "real world". Then, after they are running, we wonder why they don't have any real-world understanding. AGI will necessarily emerge in the context of robotics, as robots are the only technology based on real-world interaction.

Consider the self-driving car, which is just a big, autonomous, mobile robot. Currently being created as narrow AI, abstract concepts like "obstacle", "destination", and "pedestrian" will eventually need real-world meanings—meanings which would be impossible within the controlled verbal-only environment of Watson, for example.

Once this real-world understanding emerges in various robotic areas, it will be transferred to permeate most other areas of computation.

In Section I of this book, I'll present an overview of future intelligence in computers—contending that computers will be fast enough and that the software development is inevitable. I also introduce a plausible General Theory of Intelligence, which forms the basis of forecasts about intelligent machine behavior.

In Section II, I expand on the General Theory with a map of various observable facets of intelligence—many of which exist in today's autonomous robots. Then I'll walk through the behavior of a system with all these facets to show how it would act in an intelligent way.

In Section III, I'll predict how the future could unfold with machines based on this intelligence theory. While there are definite risks, I will show how human attitudes will mitigate or exacerbate these risks. As a future with intelligent computers is inevitable, I trust we will make the right decisions.