

The broader lesson here is that when algorithms can understand visual and perceptual information, nearly all aspects of human life will fall under algorithmic sway. This is a macroscopic trend, and it's not clear that any one engineer will have the power to prevent this future from coming into existence. Nonetheless, engineers retain the ability to vote with their feet. Your skills are valuable and in demand; don't work for companies following unethical practices and building potentially dangerous systems.



Bias in AI

Machine learning and deep learning provide the capabilities to learn interesting models from data without too much effort. This solidly mathematical process can provide the mirage of objectivity. However, it is strongly worth noting that all sorts of bias can creep into such analyses. Biases in the underlying data, drawn from historical, prejudiced records, can induce models to learn fundamentally unfair models. Google infamously once learned that a flawed visual prediction model had labeled black consumers as gorillas, likely due to biased training data that lacked adequate representation of people of color. While this system was rapidly corrected once brought to Google's notice, such failures are deeply troubling and are emblematic of more fundamental problems of exclusion in the technology industry.

As AI is increasingly used in applications such as prisoner parole granting and loan approval processes, it becomes increasingly important for us to ensure that our models aren't making racist assumptions or learning biases already present in historical data. If you are working on sensitive data, making predictions that may alter the course of human lives, check twice and check thrice to make sure that your systems aren't falling prey to biases.

Is Artificial General Intelligence Imminent?

There are widespread discussions about whether artificial general intelligence (AGI) will soon come into existence. Experts disagree strongly over whether AGI is worth seriously planning for. Our view is that while there's no harm in doing research on "AI value alignment" and "safe reward function" design, the artificial intelligence systems of today and the foreseeable future are unlikely to rapidly achieve sentience. As you will have learned first hand, most deep learning systems are simply sophisticated numerical engines, prone to many finicky numerical stability issues. It will likely take decades of fundamental advances before general intelligence becomes an issue. At the same time, as we've discussed in the previous section, artificial intelligence is already having dramatic impact on human societies and industries. It is absolutely worth

worrying about the effects of AI without the need to conjure superintelligent bogeymen.



The Superintelligent Fallacy

The book *Superintelligence* by Nick Bostrom (Oxford University Press) has had a profound impact upon the discourse surrounding AI. The basic premise of the book is that an intelligence explosion could occur when models become capable of recursively improving themselves. In itself, the premise of the book isn't that radical. If AGI were to come into existence, there's no reason to suppose that it couldn't succeed in improving itself rapidly.

At the same time, deep learning expert Andrew Ng has gone on the record stating that worrying about superintelligence is like worrying about overpopulation on Mars. One day, humanity is likely to reach Mars. When enough people land on Mars, overcrowding will likely exist and may even be a very serious problem. None of this changes the fact that Mars today is an empty wasteland. So too is the state of the literature on creating generally intelligent AI!

Now, this last statement is hyperbolic. Solid progress in reinforcement learning and generative modeling holds much promise for creating more intelligent agents. But, stressing over the possibilities for superintelligent entities detracts from the very real challenges of automation coming our way. Of course, this doesn't even mention other serious challenges facing us, such as global warming.

Where to Go from Here?

If you've read along carefully in this book and have spent effort working with our code samples in the associated GitHub repo, congrats! You have now mastered the fundamentals of practical machine learning. You will be able to train effective machine learning systems in practice.

However, machine learning is a very rapidly evolving field. The explosive growth of the field has meant that dozens of worthwhile new models are discovered each year. Practicing machine learners should constantly remain on the lookout for new models. When looking at new models, a helpful trick for evaluating their usefulness is to try to think about how you can apply the model to problems you or your organization cares about. This test provides a good way to organize the large influx of models from the research community, and will give you a tool to prioritize your learning on the techniques that really matter to you.

As a responsible machine learner, make sure to think about what your data science models are being used for. Ask yourself whether your work on machine learning is being used to improve human welfare. If the answer is no, then realize that with your