

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

skills, you have the ability to find a job where you can use your machine learning superpowers for good, not evil.

Finally, we hope that you'll have lots of fun. Deep learning is an incredibly vibrant area of human inquiry filled with exciting new discoveries, brilliant people, and the possibility of profound impact. It's been our pleasure to share our excitement and passion for the field with you, and we hope you'll pay forward our efforts by sharing your knowledge of deep learning with the world around you.

Symbols

2D convolutions, 138

A

a.eval(), 30

A3C algorithm

 A3C loss function, 196

 defining workers, 198

 overview of, 192

 training the policy, 201

A3C.fit() method, 201

accuracy, 107

accuracy_score(), 99

acknowledgments, xii

actions, 169

activations, 89

adding tensors, 32

add_output() method, 191

advantage functions, 178

adversarial models, 132

agents, 173

AI winters, 85

AlexNet, 6, 119

algorithms

 A3C algorithm, 192-196

 asynchronous training, 179

 black-box, 110

 catastrophic forgetting and, 177

 finding baselines, 111

 for reinforcement learning, 175-179

 graduate student descent, 113

 grid search, 114

 policy learning, 177

 Q-learning, 176

 random hyperparameter search, 115

AlphaGo, 12, 172, 174, 179, 196

Anaconda Python, 29, 94

architectural primitives, 3-5

architectures

 AlexNet, 6

 AlphaGo, 12

 generative adversarial networks (GANs), 13, 155

 Google's neural machine translation (Google-NMT), 9

 LeNet, 6

 neural captioning model, 8

 Neural Turing machine (NTM), 14

 one-shot models, 10

 ResNet, 7

argparse tool, 162

artificial general intelligences (AGIs), 179, 230

artificial intelligence

 cyclical development of, 85

 overhyped claims regarding, 84

ASIC (application specific integrated circuits), 209

Asynchronous Advantage Actor-Critic (see A3C algorithm)

asynchronous reinforcement, 192

asynchronous training, 179

ATARI arcade games, 169

atrous convolutions, 126

attributions, x

autoencoders, 131, 155

automated statistician project, 111

automatic differentiation, 86
 (see also backpropagation)