Suggested Readings

From “Rebooting AI: Building Artificial Intelligence We Can Trust” by Gary Marcus and Ernest Davis, September 2019.

AI in General: The leading textbook for artificial intelligence, and the most comprehensive presentation of the field as a whole, is Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach (Pearson, 2010).

A recent online series of articles, “Future of Robotics and Artificial Intelligence,” <https://rodneybrooks.com/forai-future-of-robotics-and-artificial-intelligence/> by leading roboticist Rodney Brooks (inventor of the Roomba) is very readable and very much in the spirit of our book. Brooks includes a lot of fascinating specifics, both about the practicalities of robotics, and about the history of AI.

Skepticism About AI: There have always been those with contrarian views of AI. Early works of this kind include Joseph Weizenbaum’s Computer Power and Human Reason (MIT Press, 1965); and Hubert Dreyfus, What Machines Can’t Do. The AI Delusion, by Gary Smith; Artifictional Intelligence: Against Humanity’s Surrender to Computers, by Harry Collins; and Artificial Unintelligence: How Computers Misunderstand the World, by Meredith Broussard are recent books in a similar vein.

What’s at Stake: Several important books have been published recently on the risks in AI, short and long term. Weapons of Math Destruction, by Cathy O’Neil, and Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, by Virginia Eubanks, discuss the potential for social abuse inherent in the use of big data and machine learning by government, insurance, employers, and so on.

Machine Learning and Deep Learning: The central chapters of The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, by Pedro Domingos (Basic Books, 2015), are a very readable introduction to machine-learning technologies, with chapters on each of the major approaches to machine learning. The Deep Learning Revolution, by Terrence Sejnowski, gives a historical and biographical account. Important recent textbooks in ML include Machine Learning: A Probabilistic Perspective, by Kevin Murphy, and Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. There are many free machine-learning software libraries and data sets that available online including Weka Data Mining Software[[1]](#endnote-1), Pytorch[[2]](#endnote-2), fast.ai[[3]](#endnote-3), TensorFlow[[4]](#endnote-4), Zach Lipton’s interactive Jupyter notebooks,[[5]](#endnote-5) and Andrew Ng’s popular machine-learning course on Coursera.[[6]](#endnote-6) Guides for these include Introduction to Machine Learning with Python, by Andreas Müller and Sarah Guido, and Deep Learning with Python, by François Chollet.

AI Systems That Read: There is not much written specifically for the lay reader, but the textbooks in this area often include substantial sections that are accessible to the non-expert. The standard textbooks are Speech and Language Processing, by Daniel Jurafsky and James H. Martin, and Foundations of Statistical Natural Language Processing, by Christopher Manning and Hinrich Schütze. Introduction to Information Retrieval, by Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze, is a fine introduction to web search engines and similar programs. As with machine learning, there are software libraries and data sets available online; the most widely used are the Natural Language Toolkit (usually abbreviated NLTK) at https://www.nltk.org and the Stanford Core NLP, at https://stanfordnlp.github.io/CoreNLP/; Natural Language Processing with Python: Analyzing Text with Steven Bird, Ewan Klein, and Edward Loper is a guide to using NLTK in programs. Douglas Hofstadter’s article “The Shallowness of Google Translate” (The Atlantic, Jan. 30, 2018) is an enjoyable and insightful analysis of the limitations of current approaches to machine translation.

Robotics: Other than the Rodney Brooks online articles mentioned above, there is a dearth of useful popular science writing about robotics. Matthew Mason’s fine survey article “Toward Robotic Manipulation” (2018) discusses both biological and robot manipulation. Modern Robotics: Mechanics, Planning, and Control, by Kevin Lynch and Frank Park, is an introductory textbook. Planning Algorithms, by Steven LaValle, is an overview of high-level planning for robotic motion and manipulation.

Mind: Of course, there is no end to the literature here. Particular favorites of ours include The Language Instinct and Words and Rules: The Ingredients of Language, both by Steven Pinker, for linguistics; How the Mind Works and The Stuff of Thought, by Pinker, Kluge, by Gary Marcus, and Thinking, Fast and Slow, by Daniel Kahneman for psychology; and Brainstorms, by Daniel Dennett, and Human Knowledge: Its Scope and Limits, by Bertrand Russell, for epistemology. Gary’s more technical book The Algebraic Mind, written in 2001, presages many of the issues that affect contemporary deep learning.

Commonsense Reasoning: A recent article by the authors of this book, “Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence,” is similar to chapter 7, but it is longer and includes more detail. Common Sense, the Turing Test, and the Quest for Real AI, by Hector Levesque, argues, as we have, that commonsense reasoning is a critical step in achieving genuine intelligence. Representations of Commonsense Knowledge, by Ernest Davis, is a textbook on the use of mathematical logic for representing commonsense knowledge. The Handbook of Knowledge Representation, edited by Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter, is a useful collection of surveys for more in-depth study. The Book of Why: The New Science of Cause and Effect, by Judea Pearl and Dana Mackenzie, discusses automating causal reasoning.

Trust: Moral Machines: Teaching Robots Right from Wrong by Wendell Wallach and Colin Allen, and Robot Ethics: The Ethical and Social Implications of Robotics, edited by Patrick Lin and Keith Abney, discuss the problems of instilling a moral sense into robots and AI systems.

SuperIntelligence: Paths, Dangers, Strategies, by Nick Bostrom, argues that AI will inevitably undergo a “Singularity” in which it rapidly gains intelligence and passes out of human control. Bostrom describes a variety of scenarios, ranging from dystopian to apocalyptic, for what this would mean for the human race, and discusses the possibility of developing strategies to make sure that AI remains benevolent.

Future of AI: Discussions of the long-term impact of AI on human life and society include Life 3.0: Being Human in the Age of Artificial Intelligence, by Max Tegmark; Abundance: The Future Is Better Than You Think, by Peter Diamandis and Steven Kotler; Our Final Invention: Artificial Intelligence and the End of the Human Era, by James Barrat; Artificial Intelligence: A Futuristic Approach, by Roman Yampolskiy; and The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity, by Bryon Reece. Machines That Think: The Future of Artificial Intelligence, by Toby Walsh, includes an extensive discussion of the impact of AI in the short and- long-term future, particularly its impact on employment. He also discusses the many different kinds of efforts under way, from research labs to organizations to statements of principles, to ensure that, overall, AI remains safe and beneficent.

1. Weka Data Mining Software: https://www.cs.waikato.ac.nz/ml/weka/. [↑](#endnote-ref-1)
2. Pytorch: https://pytorch.org. [↑](#endnote-ref-2)
3. fast.ai: <https://www.fast.ai/>. [↑](#endnote-ref-3)
4. TensorFlow: https://www.tensorflow.org/. [↑](#endnote-ref-4)
5. Zach Lipton’s interactive Jupyter notebooks: https://github.com/zackchase/mxnet-the-straight-dope. [↑](#endnote-ref-5)
6. Andrew Ng’s popular machine-learning course: https://www.coursera.org/learn/machine-learning. [↑](#endnote-ref-6)