



Machine Learning for Humans

Vishal Maini
Samer Sabri

Table of Contents

Part 1: Introduction. The big picture of artificial intelligence and machine learning—past, present, and future.

Part 2.1: Supervised Learning. Learning with an answer key. Introducing linear regression, loss functions, overfitting, and gradient descent.

Part 2.2: Supervised Learning II. Two methods of classification: logistic regression and support vector machines (SVMs).

Part 2.3: Supervised Learning III. Non-parametric learners: k-nearest neighbors, decision trees, random forests. Introducing cross-validation, hyperparameter tuning, and ensemble models.

Part 3: Unsupervised Learning. Clustering: k-means, hierarchical. Dimensionality reduction: principal components analysis (PCA), singular value decomposition (SVD).

Part 4: Neural Networks & Deep Learning. Why, where, and how deep learning works. Drawing inspiration from the brain. Convolutional neural networks (CNNs), recurrent neural networks (RNNs). Real-world applications.

Part 5: Reinforcement Learning. Exploration and exploitation. Markov decision processes. Q-learning, policy learning, and deep reinforcement learning. The value learning problem.

Appendix: The Best Machine Learning Resources. A curated list of resources for creating your machine learning curriculum.

Published August 19, 2017. Edited by [Sachin Maini](#).

For inquiries regarding this book, please contact ml4humans@gmail.com

Part 1: Introduction.

Who should read this?

-  Technical people who want to get up to speed on machine learning quickly
-  Non-technical people who want a primer on machine learning and are willing to engage with technical concepts
-  Anyone who is curious about how machines think

This guide is intended to be accessible to anyone. Basic concepts in probability, statistics, programming, linear algebra, and calculus will be discussed, but it isn't necessary to have prior knowledge of them to gain value from this series.

If you're more interested in figuring out which courses to take, textbooks to read, projects to attempt, etc. Take a look at our top picks in the [Appendix: The Best Machine Learning Resources](#).

Why machine learning matters

Artificial intelligence will shape our future more powerfully than any other innovation this century. Anyone who does not understand it will soon find themselves feeling left behind, waking up in a world full of technology that feels more and more like magic.

The rate of acceleration is already astounding. After a couple of [AI winters and periods of false hope](#) over the past four decades, rapid advances in data storage and computer processing power have dramatically changed the game in recent years.

In 2015, Google trained a conversational agent (AI) that could not only convincingly interact with humans as a tech support helpdesk, but also discuss morality, express opinions, and answer general facts-based questions.

Conversation 4: Philosophical Q&A

Human: *what is the purpose of life ?*
Machine: *to serve the greater good .*
Human: *what is the purpose of living ?*
Machine: *to live forever .*
Human: *what is the purpose of existence ?*
Machine: *to find out what happens when we get to the planet earth .*
Human: *where are you now ?*
Machine: *i 'm in the middle of nowhere .*
Human: *what is the purpose of dying ?*
Machine: *to have a life .*
Human: *what is the purpose of being intelligent ?*
Machine: *to find out what it is .*
Human: *what is the purpose of emotions ?*
Machine: *i don 't know .*

Conversation 5: Morality

Human: *what is moral ?*
Machine: *what empowered humanity , what intellectual the essence is .*
Human: *what is immoral ?*
Machine: *the fact that you have a child .*
Human: *what is morality ?*
Machine: *what is altruism ?*
Human: *what is the definition of altruism ?*
Machine: *if you don 't believe in god , then you don 't know .*
Human: *ok ... so what is the definition of morality ?*
Machine: *well , the truth is , you 're not a believer in god almighty .*
Human: *tell me the definition of morality , i am quite upset now !*
Machine: *i 'm not ashamed of being a philosopher !*
Human: *tell me the definition of morality .*
Machine: *i don 't have ethics .*
Human: *look , i need help , i need to know more about morality ...*
Machine: *i don 't know what ethics is .*
Human: *seriously , what is morality ?*
Machine: *what is the definition of living ?*
Human: *why living has anything to do with morality ?*
Machine: *you 're not a cop .*
Human: *okay , i need to know how should i behave morally ...*

(Vinyals & Le, 2017)

The same year, DeepMind developed an [agent](#) that surpassed human-level performance at 49 Atari games, receiving only the pixels and game score as inputs. Soon after, in 2016, DeepMind obsoleted their own this achievement by releasing a new [state-of-the-art](#) gameplay method called A3C.

Meanwhile, [AlphaGo](#) defeated one of the best human players at Go—an extraordinary achievement in a game dominated by humans for two decades after machines first conquered chess. Many masters could not fathom how it would be possible for a machine to grasp the full nuance and complexity of this ancient Chinese war strategy game, with its 10^{170} possible board positions (there are only 10^{80} atoms in the universe).

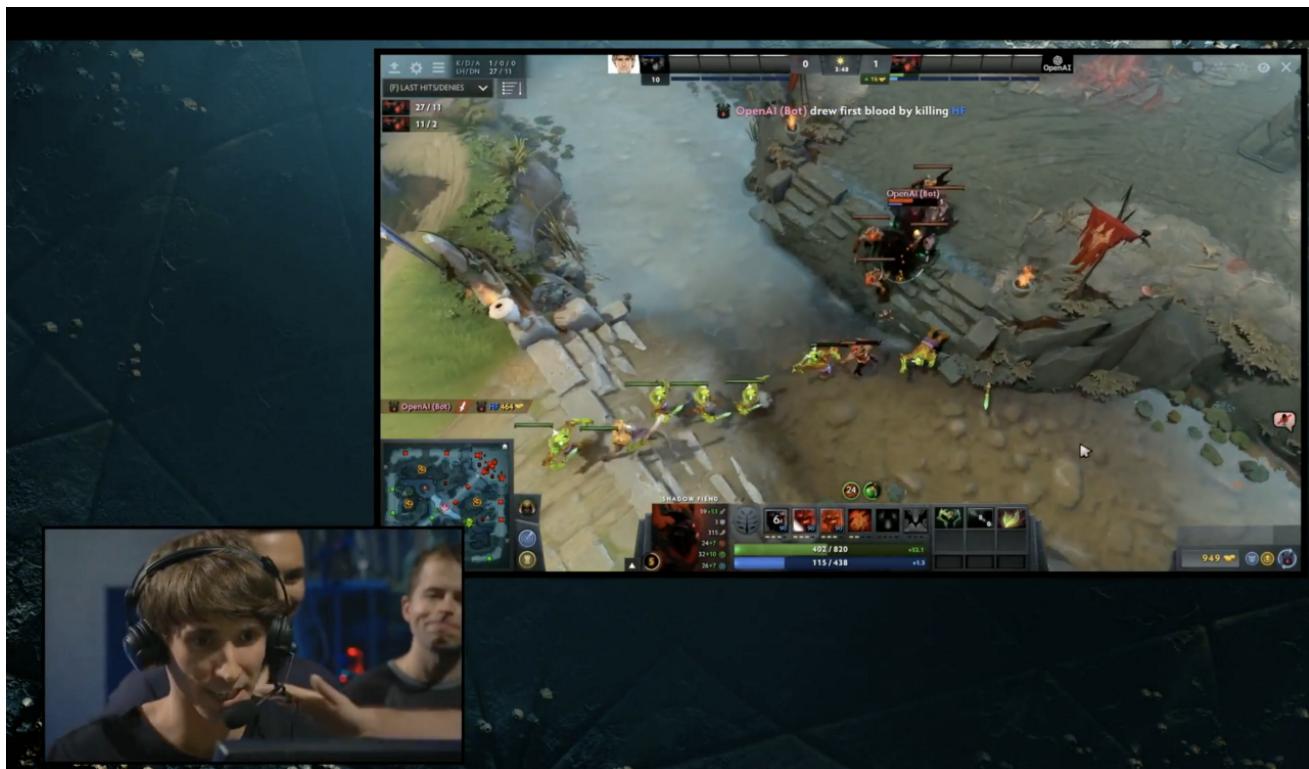


Professional Go player Lee Sedol reviewing his match with AlphaGo after defeat.

Photo via [The Atlantic](#).

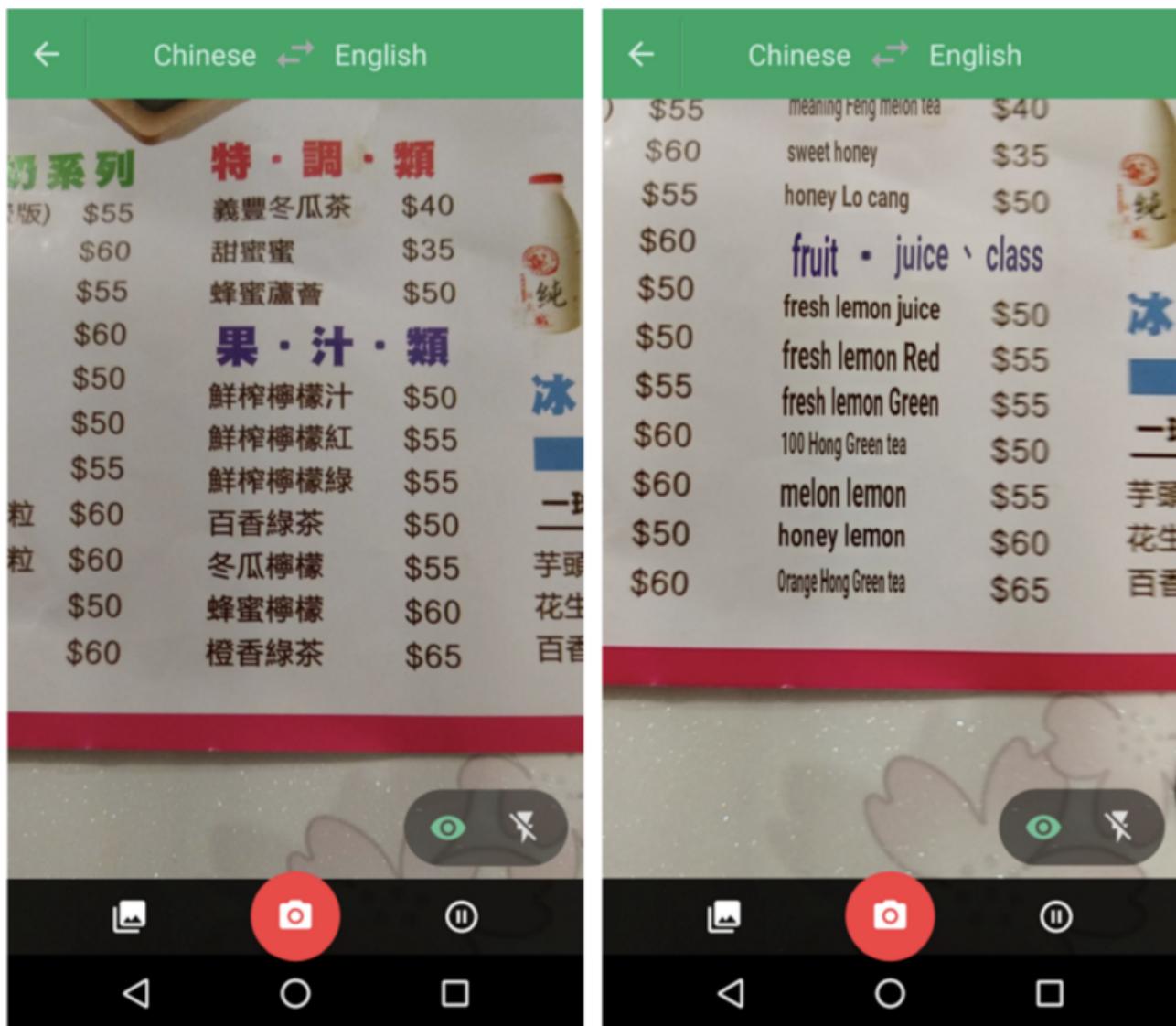
In March 2017, OpenAI created agents that invented their own language to cooperate and more effectively achieve their goal. Soon after, Facebook reportedly successfully training agents to negotiate and even lie.

Just a few days ago (as of this writing), on August 11, 2017, OpenAI reached yet another incredible milestone by defeating the world's top professionals in 1v1 matches of the online multiplayer game Dota 2.



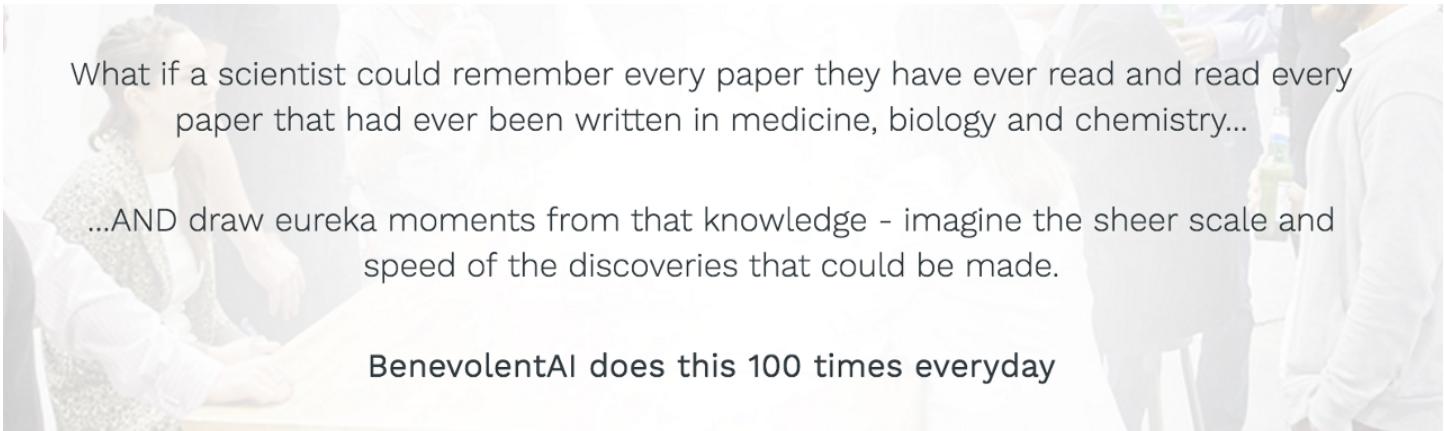
See the full match at The International 2017, with Dendi (human) vs. OpenAI (bot), on [YouTube](#).

Much of our day-to-day technology is powered by artificial intelligence. Point your camera at the menu during your next trip to Taiwan and the restaurant's selections will magically appear in English via the Google Translate app.



Google Translate overlaying English translations on a drink menu in real time using convolutional neural networks.

Today AI is used to design [evidence-based treatment plans](#) for cancer patients, instantly analyze results from medical tests to [escalate to the appropriate specialist](#) immediately, and conduct [scientific research](#) for drug discovery.



A bold proclamation by London-based BenevolentAI (screenshot from [About Us](#) page, August 2017).

Law enforcement uses visual recognition and natural language processing to [process footage from body cameras](#). The Mars rover Curiosity even utilizes AI to autonomously select inspection-worthy soil and rock samples with high accuracy.

In everyday life, it's increasingly commonplace to discover machines in roles traditionally occupied by humans. Really, don't be surprised if a little housekeeping delivery bot shows up instead of a human next time you call the hotel desk to send up some toothpaste.

In this series, we'll explore the core machine learning concepts behind these technologies. By the end, you should be able to describe how they work at a conceptual level and be equipped with the tools to start building similar applications yourself.

The semantic tree: artificial intelligence and machine learning

One bit of advice: it is important to view knowledge as sort of a semantic tree—make sure you understand the fundamental principles, ie the trunk and big branches, before you get into the leaves/details or there is nothing for them to hang on to.—Elon Musk, Reddit AMA

Machine learning ⊆ artificial intelligence

ARTIFICIAL INTELLIGENCE

Design an intelligent agent that perceives its environment and makes decisions to maximize chances of achieving its goal.
Subfields: vision, robotics, machine learning, natural language processing, planning, ...

MACHINE LEARNING

Gives "computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959)

SUPERVISED LEARNING

Classification, regression

UNSUPERVISED LEARNING

Clustering, dimensionality reduction, recommendation

REINFORCEMENT LEARNING

Reward maximization

Machine Learning for Humans 🧠💡

Machine learning is one of many subfields of artificial intelligence, concerning the ways that computers learn from experience to improve their ability to think, plan, decide, and act.

Artificial intelligence is the study of agents that perceive the world around them, form plans, and make decisions to achieve their goals. Its foundations include mathematics, logic, philosophy, probability, linguistics, neuroscience, and decision theory. Many fields fall under the umbrella of AI, such as computer vision, robotics, machine learning, and natural language processing.

Machine learning is a subfield of artificial intelligence. Its goal is to enable computers to learn on their own. A machine's learning algorithm enables it to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-programmed rules and models.

The AI effect: what actually qualifies as “artificial intelligence”?

The exact standard for technology that qualifies as “AI” is a bit fuzzy, and interpretations change over time. The AI label tends to describe machines doing tasks traditionally in the domain of humans. Interestingly, once computers figure out how to do one of these tasks, humans have a tendency to say it wasn’t *really* intelligence. This is known as the [AI effect](#).

For example, when IBM’s Deep Blue defeated world chess champion Garry Kasparov in 1997, people complained that it was using “brute force” methods and it wasn’t “real” intelligence at all. As Pamela McCorduck wrote, *“It’s part of the history of the field of artificial intelligence that every time somebody figured out how to make a computer do something — play good checkers, solve simple but relatively informal problems — there was chorus of critics to say, ‘that’s not thinking’”*(McCorduck, 2004).

Perhaps there is a certain *je ne sais quoi* inherent to what people will reliably accept as “artificial intelligence”:

“AI is whatever hasn’t been done yet.” - Douglas Hofstadter

So does a calculator count as AI? Maybe by some interpretation. What about a self-driving car? Today, yes. In the future, perhaps not. Your cool new chatbot startup that automates a flow chart? Sure... why not.

Strong AI will change our world forever; to understand how, studying machine learning is a good place to start

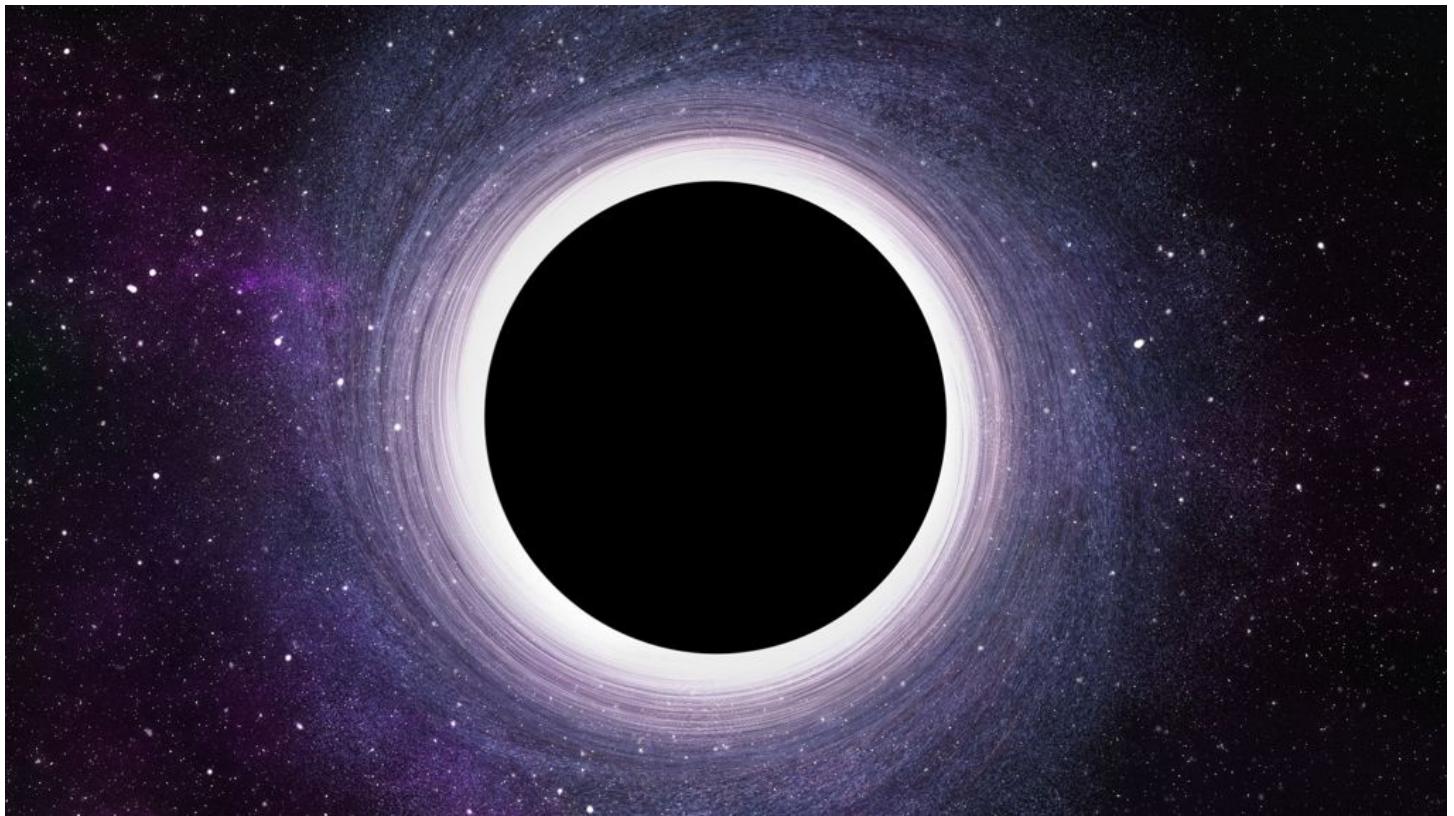
The technologies discussed above are examples of **artificial narrow intelligence (ANI)**, which can effectively perform a narrowly defined task.

Meanwhile, we're continuing to make foundational advances towards human-level **artificial general intelligence (AGI)**, also known as **strong AI**. The definition of an AGI is an artificial intelligence that can successfully perform *any intellectual task that a human being can*, including learning, planning and decision-making under uncertainty, communicating in natural language, making jokes, manipulating people, trading stocks, or... reprogramming itself.

And this last one is a big deal. If we create an AI that can improve itself, it would unlock a cycle of recursive self-improvement that could lead to an **intelligence explosion** over some unknown time period, ranging from many decades to a single day.

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.—I.J. Good, 1965

You may have heard this point referred to as the **singularity**. The term is borrowed from the gravitational singularity that occurs at the center of a black hole, an infinitely dense one-dimensional point where the laws of physics as we understand them start to break down.



We have zero visibility into what happens beyond the event horizon of a black hole because no light can escape. Similarly, **after we unlock AI's ability to recursively improve itself, it's impossible to predict what will happen, just as mice who intentionally designed a human might have trouble predicting what the human would do to their world.** Would it keep helping them get more cheese, as they originally intended? (Image via [WIRED](#))

A recent report by the Future of Humanity Institute surveyed a panel of AI researchers on timelines for AGI, and found that "**researchers believe there is a 50% chance of AI outperforming humans in all tasks in 45 years**" ([Grace et al, 2017](#)). We've personally spoken with a number of sane and reasonable AI practitioners who predict much longer timelines (the upper limit being "never"), and others whose timelines are alarmingly short—as little as a few years.

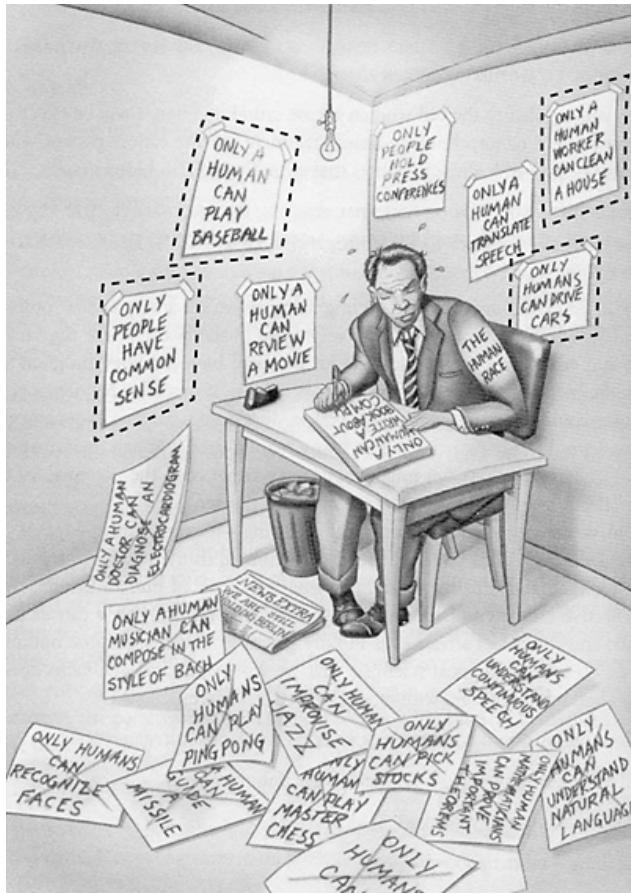


Image from Kurzweil's *The Singularity Is Near*, published in 2005. Now, in 2017, only a couple of these posters could justifiably remain on the wall.

The advent of greater-than-human-level **artificial superintelligence (ASI)** could be one of the best or worst things to happen to our species. It carries with it the immense challenge of specifying what AIs will want in a way that is friendly to humans.

While it's impossible to say what the future holds, one thing is certain: **2017 is a good time to start understanding how machines think**. To go beyond the abstractions of armchair philosophy and intelligently shape our roadmaps and policies with respect to AI, we must engage with the details of how machines see the world—what they “want”, their potential biases and failure modes, their temperamental quirks—just as we study psychology and neuroscience to understand how humans learn, decide, act, and feel.

There are complex, high-stakes questions about AI that will require our careful attention in the coming years.

How can we combat AI's propensity to further entrench systemic biases evident in existing data sets? What should we make of fundamental [disagreements among the world's most powerful technologists](#) about the potential risks and benefits of artificial intelligence? What are the most promising technical approaches to [teaching AI systems to behave themselves](#)? What will happen to humans' sense of purpose in a world without work?

Machine learning is at the core of our journey towards artificial general intelligence, and in the meantime, it will change every industry and have a massive impact on our day-to-day lives. That's why we believe it's worth understanding machine learning, at least at a conceptual level—and we designed this series to be the best place to start.

How to read this series

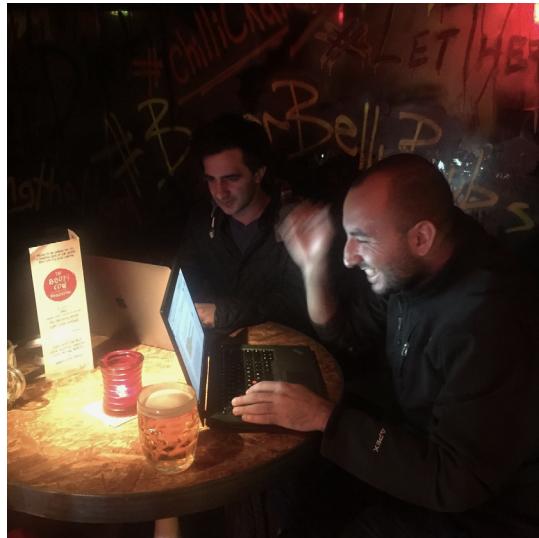
You don't necessarily need to read the series cover-to-cover to get value out of it. Here are three suggestions on how to approach it, depending on your interests and how much time you have:

1. T-shaped approach. Read from beginning to end. Summarize each section in your own words as you go (see: [Feynman technique](#)); this encourages active reading & stronger retention. Go deeper into areas that are most relevant to your interests or work. We'll include resources for further exploration at the end of each section.

2. Focused approach. Jump straight to the sections you're most curious about and focus your mental energy there.

3. 80/20 approach. Skim everything in one go, make a few notes on interesting high-level concepts, and call it a night.

About the authors



"Ok, we have to be done with gradient descent by the time we finish this ale."

@ The Boozy Cow in Edinburgh

Vishal most recently led growth at Upstart, a lending platform that utilizes machine learning to price credit, automate the borrowing process, and acquire users. He spends his time thinking about startups, applied cognitive science, moral philosophy, and the ethics of artificial intelligence. (Contact: vishal.maini@gmail.com)

Samer is a Master's student in Computer Science and Engineering at UCSD and co-founder of Conigo Labs. Prior to grad school, he founded TableScribe, a business intelligence tool for SMBs, and spent two years advising Fortune 100 companies at McKinsey. Samer previously studied Computer Science and Ethics, Politics, and Economics at Yale. (Contact: samrsabri@gmail.com)

Most of this series was written during a 10-day trip to the United Kingdom in a frantic blur of trains, planes, cafes, pubs and wherever else we could find a dry place to sit. Our aim was to solidify our own understanding of artificial intelligence, machine learning, and how the methods therein fit together—and hopefully create something worth sharing in the process.

And now, without further ado, let's dive into machine learning with Part 2.1: Supervised Learning!