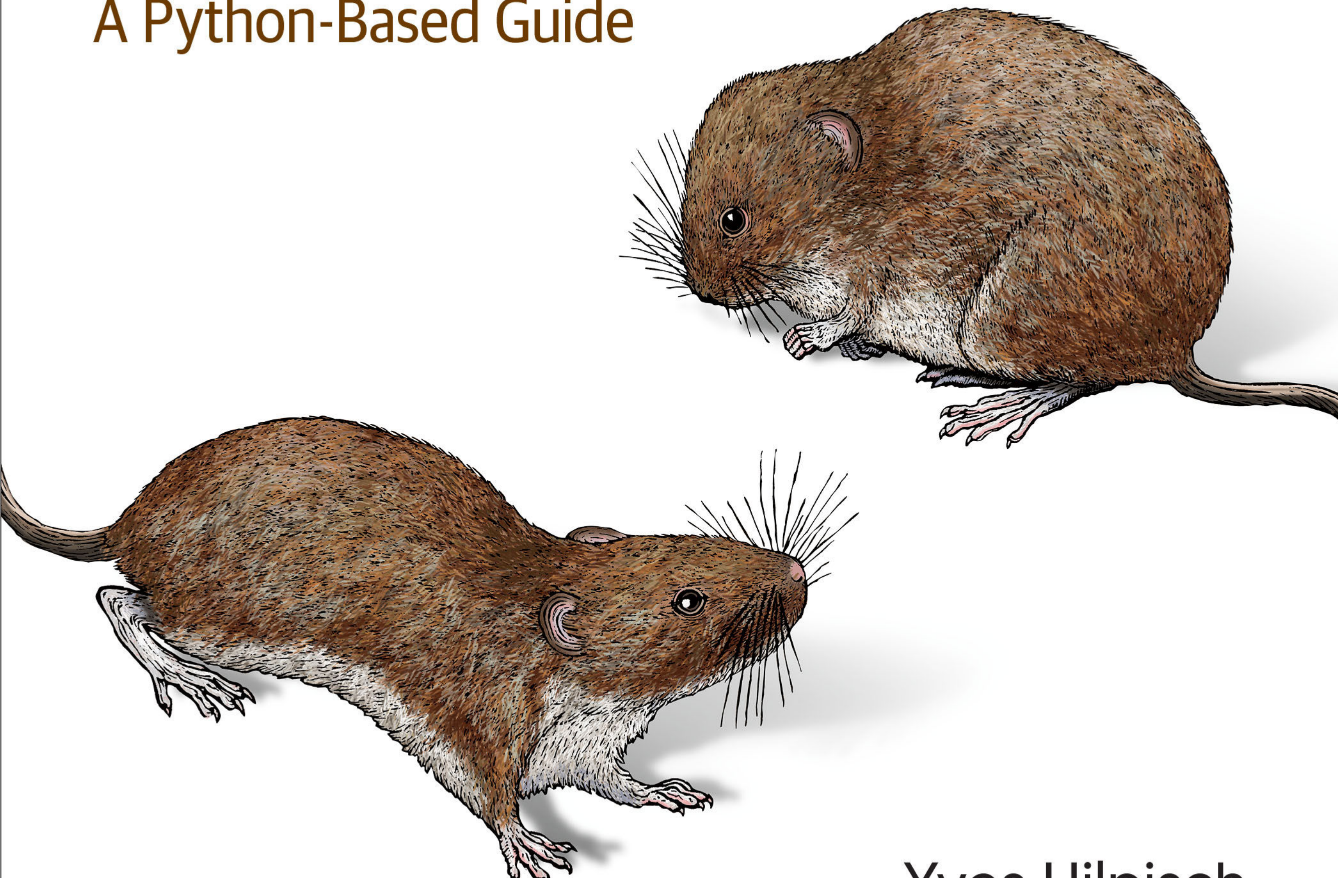
O'REILLY

Artificial Intelligence in Finance

A Python-Based Guide



Yves Hilpisch

Artificial Intelligence in Finance

A Python-Based Guide

Yves Hilpisch

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

Table of Contents

Part II. Finance and Machine Learning

Part V. Outlook

Part VI. Appendices

Preface

> Will alpha eventually go to zero for every imaginable investment strategy? More fundamentally, is the day approaching when, thanks to so many smart people and smarter computers, financial markets really do become perfect, and we can just sit back, relax, and assume that all assets are priced correctly?
>
> —Robert Shiller (2015)

Artificial intelligence (AI) rose to become a key technology in the 2010s and is assumed to be the dominating technology in the 2020s. Spurred by technological innovations, algorithmic breakthroughs, availability of big data, and ever-increasing compute power, many industries are undergoing fundamental changes driven by AI.

While media and public attention mostly focus on breakthroughs in areas such as gaming and self-driving cars, AI has also become a major technological force in the financial industry. However, it is safe to say that AI in finance is still at a nascent stage —as compared, for example, to industries such as web search or social media.

This book sets out to cover a number of important aspects related to AI in finance. AI in finance is already a vast topic, and a single book needs to focus on selected aspects. Therefore, this book covers the basics first (see Part I and Part II). It then zooms in on discovering *statistical inefficiencies* in financial markets by the use of AI and, more specifically, neural networks (see Part III). Such inefficiencies—embodied by AI algorithms that predict successfully future market movements—are a prerequisite for the exploitation of *economic inefficiencies* through algorithmic trading (see Part IV). Being able to systematically exploit statistical and economic inefficiencies would prove contradictory to one of the established theories and cornerstones in finance: the *efficient markets hypothesis* (EMH). The design of a successful *trading bot* can be considered the holy grail in finance to which AI might lead the way. This book concludes by discussing consequences of AI for the financial industry and the possibility of a financial singularity (see Part V). There is also a technical appendix that builds neural networks from scratch based on plain Python code and provides additional examples for their application (see Part VI).

The problem of applying AI to finance is not too dissimilar to other fields. Some major breakthroughs in AI in the 2010s were made possible by the application of reinforcement learning (RL) to playing arcade games, such as those from Atari published in the 1980s (see Mnih et al. 2013), and to board games, such as chess or Go (see Silver et al. 2016). Lessons learned from applying RL in gaming contexts, among other areas, are today applied to such challenging problems as designing and building autonomous vehicles or improving medical diagnostics. Table P-1 compares the application of AI and RL in different domains.

Table P-1. Comparison of AI in different domains

Domain	Agent	Goal	Approach	Reward	Obstacle	Risks
Arcade games	AI agent (software)	Maximizing game score	RL in virtual gaming environment	Points and scores	Planning and delayed rewards	None
Autonomous driving	Self-driving car (software + car)	Safely driving from location A to B	RL in virtual (gaming) environment, real-world test drives	Punishment for mistakes	Transition from virtual to physical world	Damaging property, harming people
Financial trading	Trading bot (software)	Maximizing long-term performance	RL in virtual trading environment	Financial returns	Efficient markets and competition	Financial losses

The beauty of training AI agents to play arcade games lies in the availability of a perfect virtual learning environment[1] and the absence of any kind of risk. With autonomous vehicles, the major problem arises when transitioning from virtual learning environments—for example, a computer game such as Grand Theft Auto (GTA)—to the physical world with a self-driving car navigating real streets populated by other cars and people. This leads to serious risks such as a car causing accidents or harming people.

For a trading bot, RL can also be completely virtual, that is in a simulated financial market environment. The major risks that arise from malfunctioning trading bots are financial losses and on an aggregated level, potential systematic risks due to herding by trading bots. Overall, however, the financial domain seems like an ideal place to train, test, and deploy AI algorithms.

Given the rapid developments in the field, it should even be possible for an interested and ambitious student, equipped with a notebook and Internet access, to successfully apply AI in a financial trading context. Beyond hardware and software improvements over recent years, this is due primarily to the rise of online brokers that supply histor-

1 See the Arcade Learning Environment.

ical and real-time financial data and that allow the execution of financial trades via programmatic APIs.

The book is structured in the following six parts.

Part I
: The first part discusses central notions and algorithms of AI in general, such as supervised learning and neural networks (see Chapter 1). It also discusses the concept of superintelligence, which relates to an AI agent that possesses human-level intelligence and, in some domains, superhuman-level intelligence (see Chapter 2). Not every researcher in AI believes that superintelligence is possible in the foreseeable future. However, the discussion of this idea provides a valuable framework for discussing AI in general and AI for finance in particular.

Part II
: The second part consists of four chapters and is about traditional, normative finance theory (see Chapter 3) and how the field is transformed by data-driven finance (see Chapter 4) and machine learning (ML) (see Chapter 5). Taken together, data-driven finance and ML give rise to a model-free, AI-first approach to finance, as discussed in Chapter 6.

Part III
: The third part is about discovering statistical inefficiencies in financial markets by applying deep learning, neural networks, and reinforcement learning. The part covers dense neural networks (DNNs, see Chapter 7), recurrent neural networks (RNNs, see Chapter 8), and algorithms from reinforcement learning (RL, see Chapter 9) that in turn often rely on DNNs to represent and approximate the optimal policy of the AI agent.

Part IV
: The fourth part discusses how to exploit statistical inefficiencies through algorithmic trading. Topics are vectorized backtesting (see Chapter 10), event-based backtesting and risk management (see Chapter 11), as well as execution and deployment of AI-powered algorithmic trading strategies (see Chapter 12).

Part V
: The fifth part is about the consequences that arise from AI-based competition in the financial industry (see Chapter 13). It also discusses the possibility of a financial singularity, a point in time from which AI agents dominate all aspects of finance as we know it. The discussion in this context focuses on artificial financial intelligences as trading bots that consistently generate trading profits above any human or institutional benchmark (see Chapter 14).

Part VI

The Appendix contains Python code for interactive neural network training (see Appendix A), classes for simple and shallow neural networks that are implemented from scratch based on plain Python code (see Appendix B), and an example of how to use convolutional neural networks (CNNs) for financial time series prediction (see Appendix C).

Author's Note

The application of AI to financial trading is still a nascent field, although there are a number of other books available that cover this topic at least to some extent at the time of writing. Many of these publications, however, fail to show what it entails to *economically* exploit statistical inefficiencies.

Some hedge funds already claim to exclusively rely on machine learning to manage their investors' capital. A prominent example is The Voleon Group, a hedge fund that reported more than 6 billion USD in assets under management at the end of 2019 (see Lee and Karsh 2020). The difficulty of relying on machine learning to outsmart the financial markets is reflected in the fund's performance of 7% for 2019, a year during which the S&P 500 stock index rose by almost 30%.

This book is based on years of practical experience in developing, backtesting, and deploying AI-powered algorithmic trading strategies. The approaches and examples presented are mostly based on my own research since the field is, by nature, not only nascent, but also rather secretive. The exposition and the style throughout this book are relentlessly practical, and in many instances the concrete examples are lacking proper theoretical support and/or comprehensive empirical evidence. This book even presents some applications and examples that might be vehemently criticized by experts in finance and/or machine learning.

For example, some experts in machine and deep learning, such as François Chollet (2017), outright doubt that prediction in financial markets is possible. Certain experts in finance, such as Robert Shiller (2015), doubt that there will ever be something like a financial singularity. Others active at the intersection of the two domains, such as Marcos López de Prado (2018), argue that the use of machine learning for financial trading and investing requires an industrial scale effort with large teams and huge budgets.

This book does not try to provide a balanced view of or a comprehensive set of references for all the topics covered. The presentation is rather driven by the personal opinions and experiences of the author, as well as practical considerations when providing concrete examples and Python code. Many of the examples are also chosen and tweaked to drive home certain points or to show encouraging results. Therefore, it can for sure be argued that results from many examples presented in the book suffer from data snooping and overfitting (for a discussion of these topics see Hilpisch 2020, ch. 4).

The major goal of this book is to empower the reader to use the code examples in the book as a framework to explore the exciting space of AI applied to financial trading. To achieve this goal, the book relies throughout on a number of simplifying assumptions and primarily on *financial time series data and features derived directly from such data*. In practical applications, a restriction to financial time series data is of course not necessary—a great variety of other types of data and data sources could be used as well. This book's approach to deriving features implicitly assumes that financial time series and features derived from them show *patterns* that, at least to some extent, persist over time and that can be used to predict the direction of future price movements.

Against this background, all examples and the code presented in this book are technical and illustrative in nature and do not represent any recommendation or investment advice.

For those who want to deploy approaches and algorithmic trading strategies presented in this book, my book *Python for Algorithmic Trading: From Idea to Cloud Deployment* (O'Reilly) provides more process-oriented and technical details. The two books complement each other in many respects. For readers who are just getting started with Python for finance or who are seeking a refresher and reference manual, my book *Python for Finance: Mastering Data-Driven Finance* (O'Reilly) covers a comprehensive set of important topics and fundamental skills in Python as applied to the financial domain.

References

Papers and books cited in the preface:

Chollet, Francois. 2017. *Deep Learning with Python*. Shelter Island: Manning.

Hilpisch, Yves. 2018. *Python for Finance: Mastering Data-Driven Finance*. 2nd ed. Sebastopol: O'Reilly.

———. 2021. *Python for Algorithmic Trading: From Idea to Cloud Deployment*. Sebastopol: O'Reilly.

Lee, Justina and Melissa Karsh. 2020. "Machine-Learning Hedge Fund Voleon Group Returns 7% in 2019." *Bloomberg* (January): 21 .

López de Prado, Marcos. 2018. *Advances in Financial Machine Learning*. Hoboken, NJ: John Wiley & Sons.

Mnih, Volodymyr et al. 2013. "Playing Atari with Deep Reinforcement Learning." *https://oreil.ly/-pW-1*.

Shiller, Robert. 2015. "The Mirage of the Financial Singularity." *https://oreil.ly/VRkP3*.

Silver, David et al. 2016. "Mastering the game of Go with Deep Neural Networks and Tree Search." *Nature* 529 (January): 484-489.

PART I
Machine Intelligence

> Today's algorithmic trading programs are relatively simple and make only limited use of AI. This is sure to change.
>
> —Murray Shanahan (2014)

This part is about artificial intelligence (AI) in general: *artificial* in the sense that the intelligence is not displayed by a biological organism but rather by a machine, and *intelligence* in the sense as defined by AI researcher Max Tegmark as the "ability to accomplish complex goals." This part introduces central notions and algorithms from the AI field, gives examples of major recent breakthroughs, and discusses aspects of superintelligence. It consists of two chapters:

- Chapter 1 introduces general notions, ideas, and definitions from the field of AI. It also provides several Python examples of how different algorithms can be applied in practice.
- Chapter 2 discusses concepts and topics related to artificial general intelligence (AGI) and superintelligence (SI). These types of intelligence relate to AI agents that have reached at least human level intelligence in all domains and super-human intelligence in certain domains.

About the Author

Dr. Yves J. Hilpisch is founder and managing partner of The Python Quants, a group focusing on the use of open source technologies for financial data science, artificial intelligence, algorithmic trading, and computational finance. He is also founder and CEO of The AI Machine, a company focused on AI-powered algorithmic trading via a proprietary strategy execution platform.

In addition to this book, he is the author of the following books:

- Python for Algorithmic Trading (O'Reilly, forthcoming),
- Python for Finance (2nd ed., O'Reilly, 2018),
- Derivatives Analytics with Python (Wiley, 2015), and
- Listed Volatility and Variance Derivatives (Wiley, 2017).

Yves is an adjunct professor of computational finance and lectures on algorithmic trading at the CQF Program. He is also the director of the first online training programs leading to university certificates in Python for Algorithmic Trading and Python for Computational Finance.

Yves wrote the financial analytics library DX Analytics and organizes meetups, conferences, and bootcamps about Python for quantitative finance and algorithmic trading in London, Frankfurt, Berlin, Paris, and New York. He has given keynote speeches at technology conferences in the United States, Europe, and Asia.

Colophon

The animal on the cover of *Artificial Intelligence in Finance* is a bank vole (*myodes glareolus*). These voles can be found in forests, banks, and swamps throughout Europe and Central Asia, with notable populations in Finland and the United Kingdom.

Bank voles are small, only 10-11cm in length and 17-20g on average, and have small eyes and ears. Their fur is thick and typically brown or gray, and covers their whole body. Bank voles have short tails and small brains relative to their body size. Pups are born blind and helpless in litters of four to eight, after after which they mature rather quickly, with females reaching maturity in two to three weeks and males maturing at six to eight weeks. The average lifespan for a bank vole mirrors this quick maturation, with most individuals living a half to two years.

These small rodents are primarily active during twilight, though they can be diurnal or nocturnal as well. They have an omnivorous diet consisting mostly of plant matter, and what they eat changes with the seasons. Socially, female bank voles are dominant over males, who disperse once reaching maturity, while female bank voles will typically stay closer to where they were born.

Given their relatively healthy population numbers and wide distributions, the bank vole's current conservation status is that of "Least Concern." Many of the animals on O'Reilly covers are endangered; all of them are important to the world. To learn more about how you can help, go to animals.oreilly.com.

Color illustration by Karen Montgomery, based on a black and white engraving from *British Quadrupeds*. The cover fonts are Gilroy Semibold and Guardian Sans. The text font is Adobe Minion Pro; the heading font is Adobe Myriad Condensed; and the code font is Dalton Maag's Ubuntu Mono.