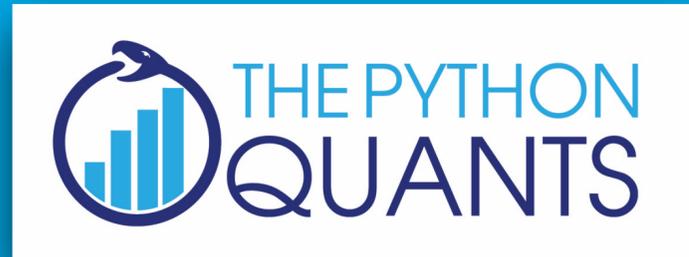


# AI-Powered Algorithmic Trading

Dr. Yves J. Hilpisch

Refinitiv Developer Day



**“Data is the new oil.”**

**“Software is eating the world.”**

**“Data and AI are eating finance.”**

# The Group



## SERVICES

for financial institutions globally



## EVENTS

for Python quants & algorithmic traders



## TRAINING

about Python for finance & algorithmic trading



## CERTIFICATION

in cooperation with university



## BOOKS

about Python and finance



## PLATFORM

for browser-based data analytics



## OPEN SOURCE

Python library for financial analytics



16 week program

150+ hours  
of instruction

**PROGRAM DIRECTOR**

Dr. Yves J. Hilpisch is founder and managing partner of The Python Quants (<http://tpq.io>), a group focusing on the use of open source technologies for financial data science, algorithmic trading and computational finance. He is the author of the books:

- He is the author of the books:
  - Python for Finance (O'Reilly)
  - Derivatives Analytics with Python (Wiley)
  - Listed Volatility and Variance Derivatives (Wiley)

He has written the financial analytics library DX Analytics (<http://dx-analytics.com>) and organizes conferences and Meetup events about Python for finance and algorithmic trading in Frankfurt, London and New York. He has given keynote speeches at technology conferences in the United States, Europe and Asia.



**UNIVERSITY CERTIFICATE  
IN PYTHON FOR  
ALGORITHMIC TRADING**



5,000+ lines  
of code

1,200 pages PDF

The Python Quants GmbH  
66333 Voelklingen  
Germany  
T/F +49 3212 112 91 94  
<http://training.tpq.io>  
[training@tpq.io](mailto:training@tpq.io)

April 2017

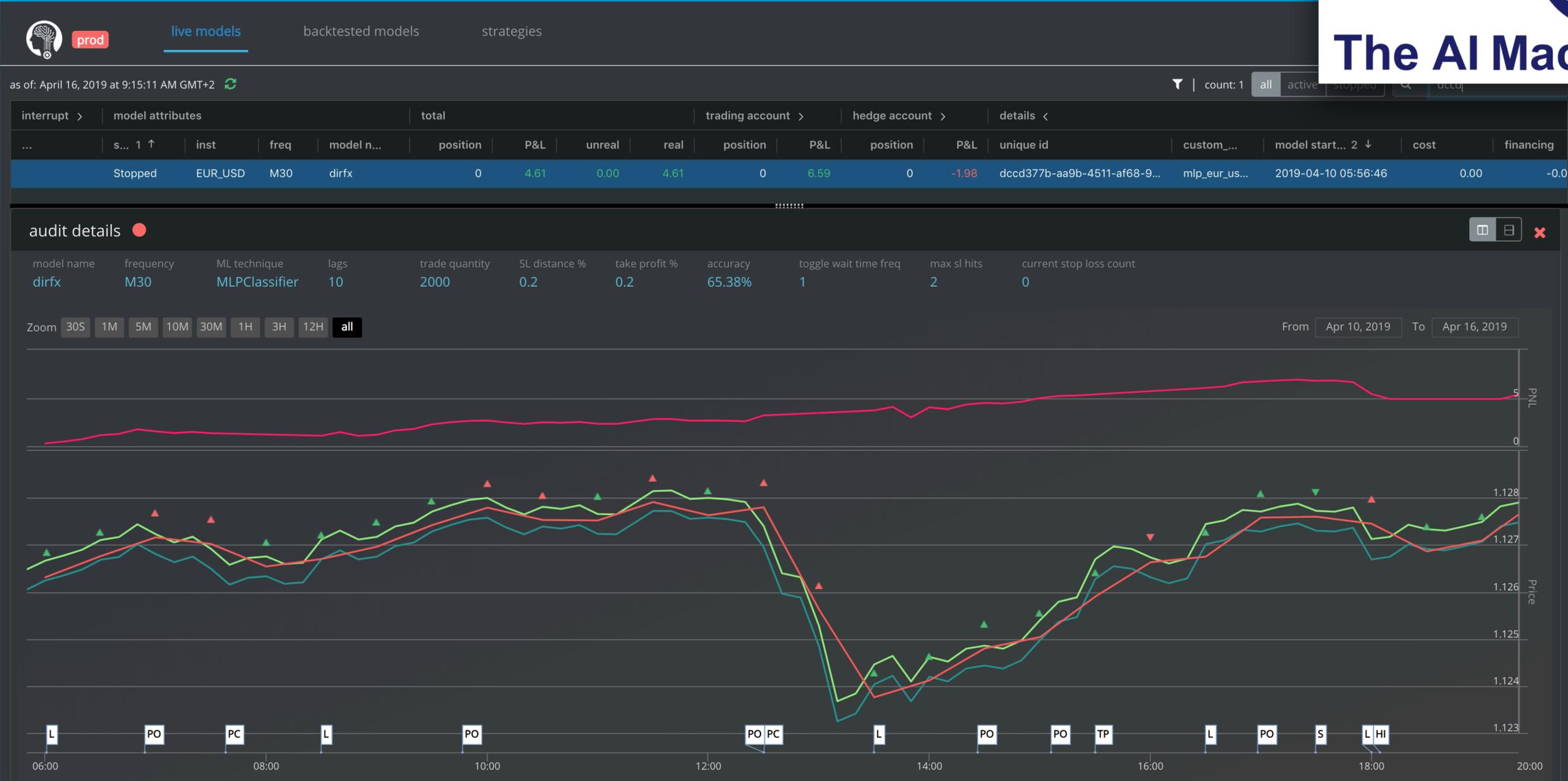
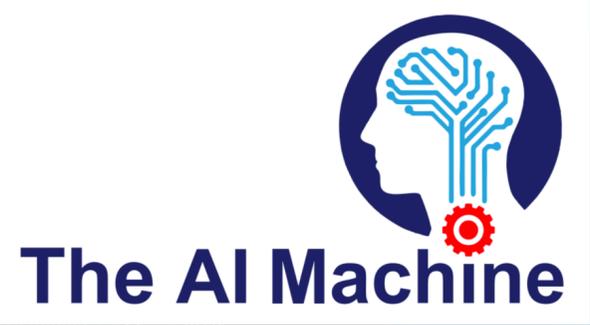
<http://certificate.tpq.io>

# Wanted: Finance graduates with AI and programming skills

Quant funds are crying out for tech skills, and schools are responding with specialist courses



Looking ahead: Pierre Foret © Magali Delporte





recognized by **Capital Markets**  
**CIO** magazine as  
Outlook

TOP 10  
**ALGO TRADING**  
SOLUTION PROVIDERS - 2019

*An annual listing of 10 companies that are at the forefront  
of providing Algo Trading solutions*

[http://certificate.tpq.io/tpq\\_top\\_algo\\_2019.pdf](http://certificate.tpq.io/tpq_top_algo_2019.pdf)

Capital Markets  
**CIO** TOP 10  
Outlook **ALGO TRADING**  
SOLUTION PROVIDERS - 2019

## The Python Quants First University Certificate in Python for Algorithmic Trading

**P**ython programming has become a key skill in the financial industry. In areas such as financial data science, computational finance or algorithmic trading, Python has established itself as the primary technological platform. At the same time, the level of Python sophistication the industry is expecting from its employees and applicants is increasing steadily. The Python Quants Group is one of the leading providers of Python for Finance training programs.

Among others, The Python Quants have tailored a comprehensive online training program leading to the first University Certificate in Python for Algorithmic Trading. Be it an ambitious student with intrigue for algorithmic trading, or a major financial institution, The Python Quants, through this systematic training program, is equipping delegates with requisite skills and tools to formulate, backtest and deploy algorithmic trading strategies based on Python.

The topics covered in the training programs offered by The Python Quants are generally not found in the typical curriculum of financial engineering or quantitative finance Master programs. Dr. Yves Hilpisch, the firm's founder and managing partner, explains, "There are courses out there that show students how to apply machine learning for the formulation and backtesting of algorithmic trading strategies. However, none of them explains the difficulties or the skills

required in deploying such algorithmic trading strategies in the real world. Besides providing an introductory course that teaches Python and financial concepts from scratch, we train our delegates and clients on how best to deploy algorithmic trading strategies in automated fashion in the cloud, with, among others, real-time risk management and monitoring," explains Hilpisch, an author of three books on

the topic, with "Python for Finance" (2nd ed., O'Reilly) being the standard reference in the field.

The organization's "Python for Algorithmic Trading University Certificate" consists of 200 hours of instruction, 1,200 pages of documentation and 1,000s of lines of Python code. In addition to offering both online and offline Python training, Hilpisch and his team also organize bespoke training events for financial institutions, hedge funds, banks, and asset management companies. "Most of the training is online since we have students and delegates from about 65 different countries in general. Most recently, we noticed that it's not just financial firms and students who want to deepen their algorithmic trading knowledge, but even professors of finance who want to get more involved in this popular topic," says Hilpisch.

While the Quant Platform is the most popular choice, especially for users in the financial sector who don't have access to a full-fledged, interactive, financial analytics environment, the team at The Python Quants is currently developing The AI Machine—a new platform which leverages artificial intelligence to formulate and deploy algorithmic trading strategies in a standardized manner. Hilpisch explains that it's relatively easy to write Python code for an algorithmic trading strategy, but the same can't be said about the deployment of such a strategy. "There are a few platforms out there that allow the formulation and backtesting of algorithmic trading strategies by the use of Python code. However, they usually stop exactly there. With The AI Machine, it is a single click on the 'GO LIVE' button and the strategy is deployed in real-time—without any changes to the strategy code itself," adds Hilpisch.

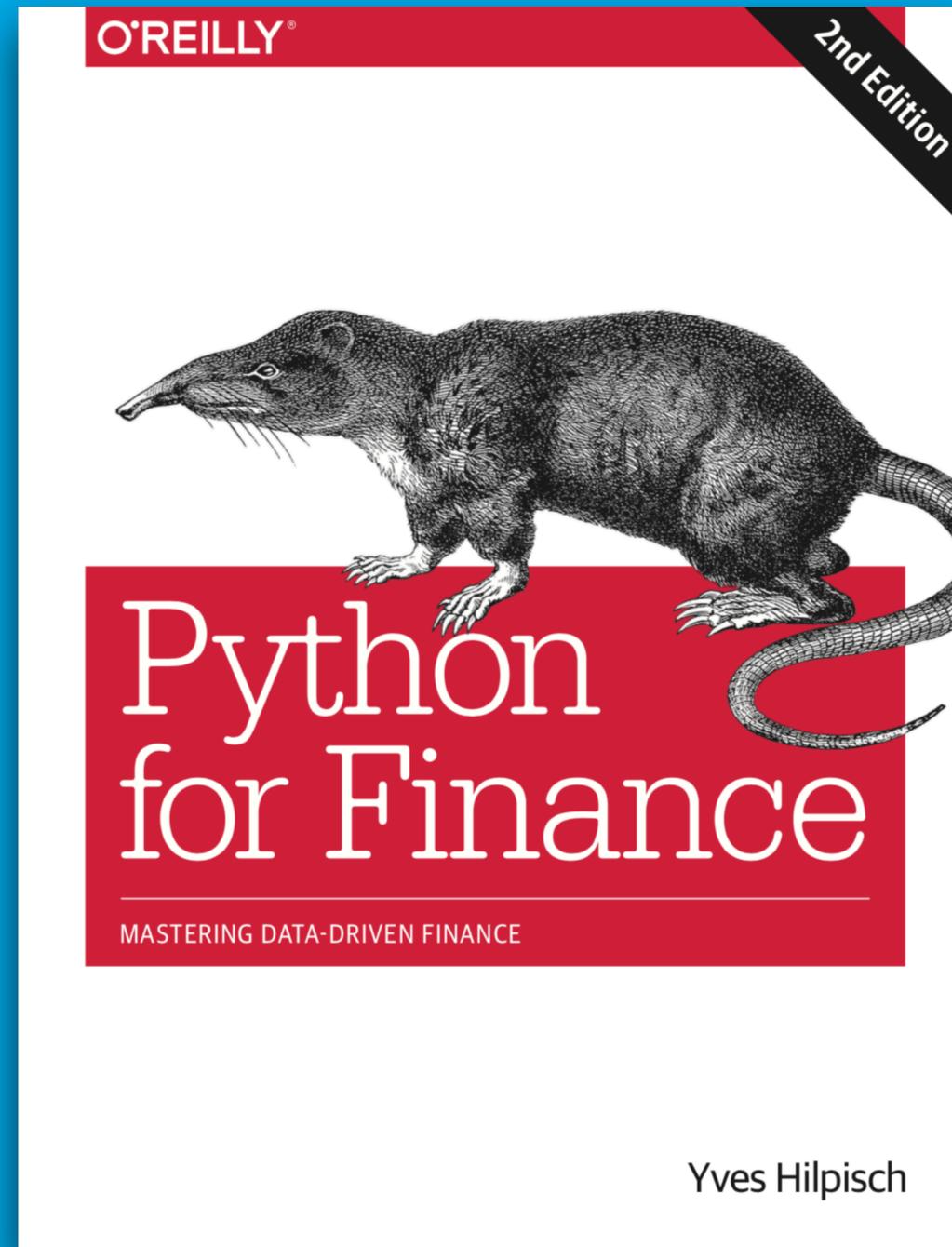
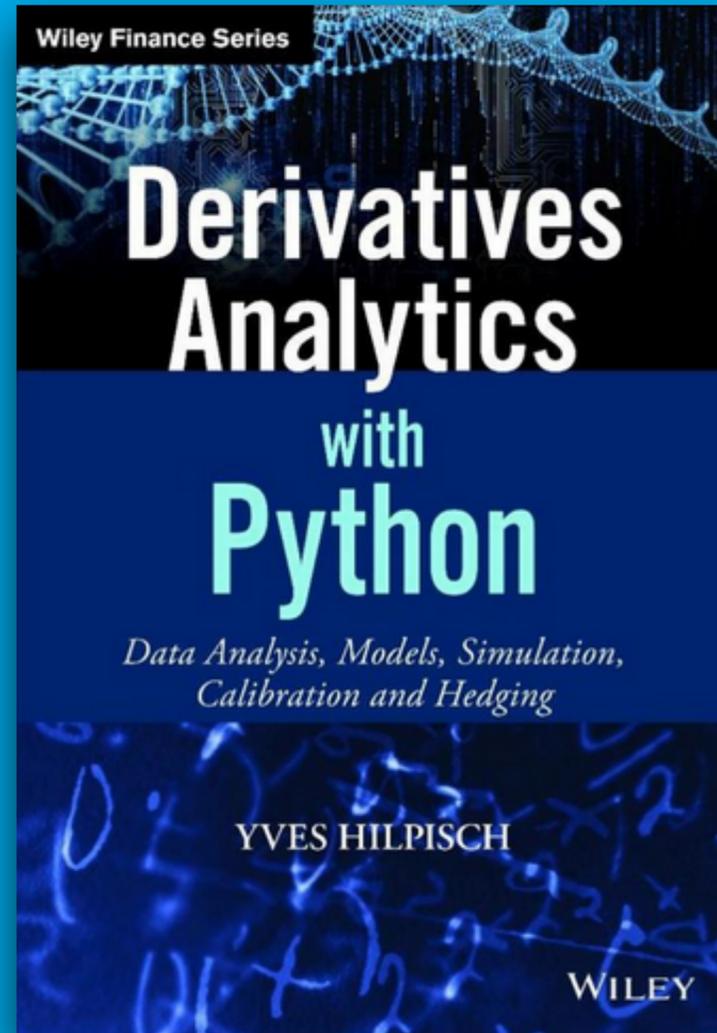
In 2019, The Python Quants will be introducing a new university certificate titled "Python for Computational Finance," which will focus more on original quantitative finance topics, such as option pricing, Monte Carlo simulation, and hedging. As financial institutions begin to perceive Python-based analytics as a prerequisite skill, the organization will continue to provide an "efficient and structured way of mastering all the tools and skills required in Python for Financial Data Science, Algorithmic Trading, and Computational Finance." **CM**



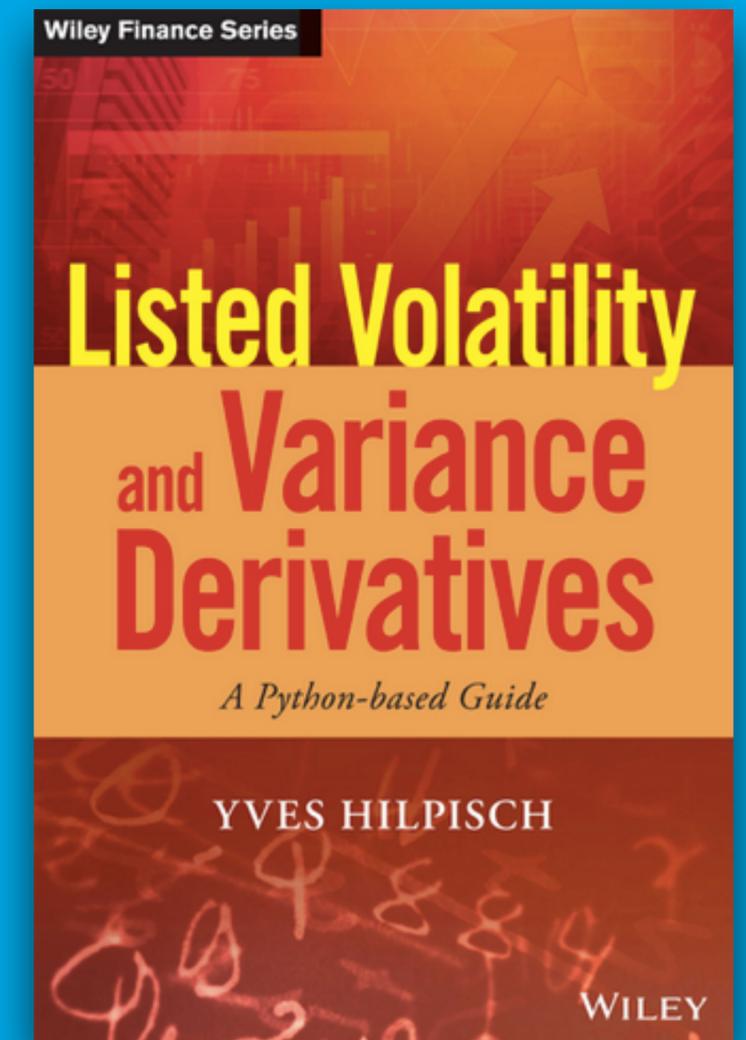
Dr. Yves Hilpisch

# About Myself

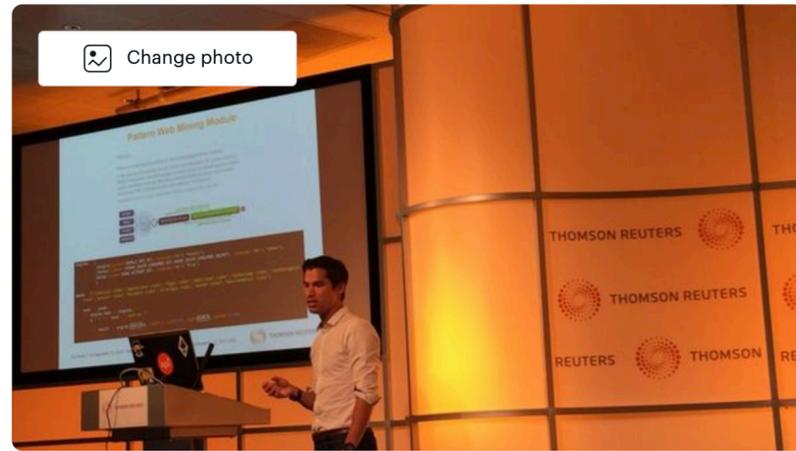




NEW book project:  
**Artificial Intelligence in Finance**  
— A Python-based Guide



<http://books.tpq.io>



Change photo

## Python for Quant Finance

London, United Kingdom  
2,817 members · Public group  
Organized by Yves Hilpisch and 2 others

Share:

About Events Members Photos Discussions More

Manage group

### What we're about

This group is about the use of Python for Quantitative Financial Applications and Interactive Financial Analytics.

### Organizers

Yves Hilpisch and others  
[Message](#)



Change photo

## Artificial Intelligence in Finance & Algorithmic Trading

New York, NY  
345 members · Public group  
Organized by Yves Hilpisch

Share:

About Events Members Photos Discussions More

Manage group

Create event

### What we're about

This Meetup group is concerned with data-driven and AI-first finance in general and algorithmic trading in particular. Its events cover the latest...

### Organizer

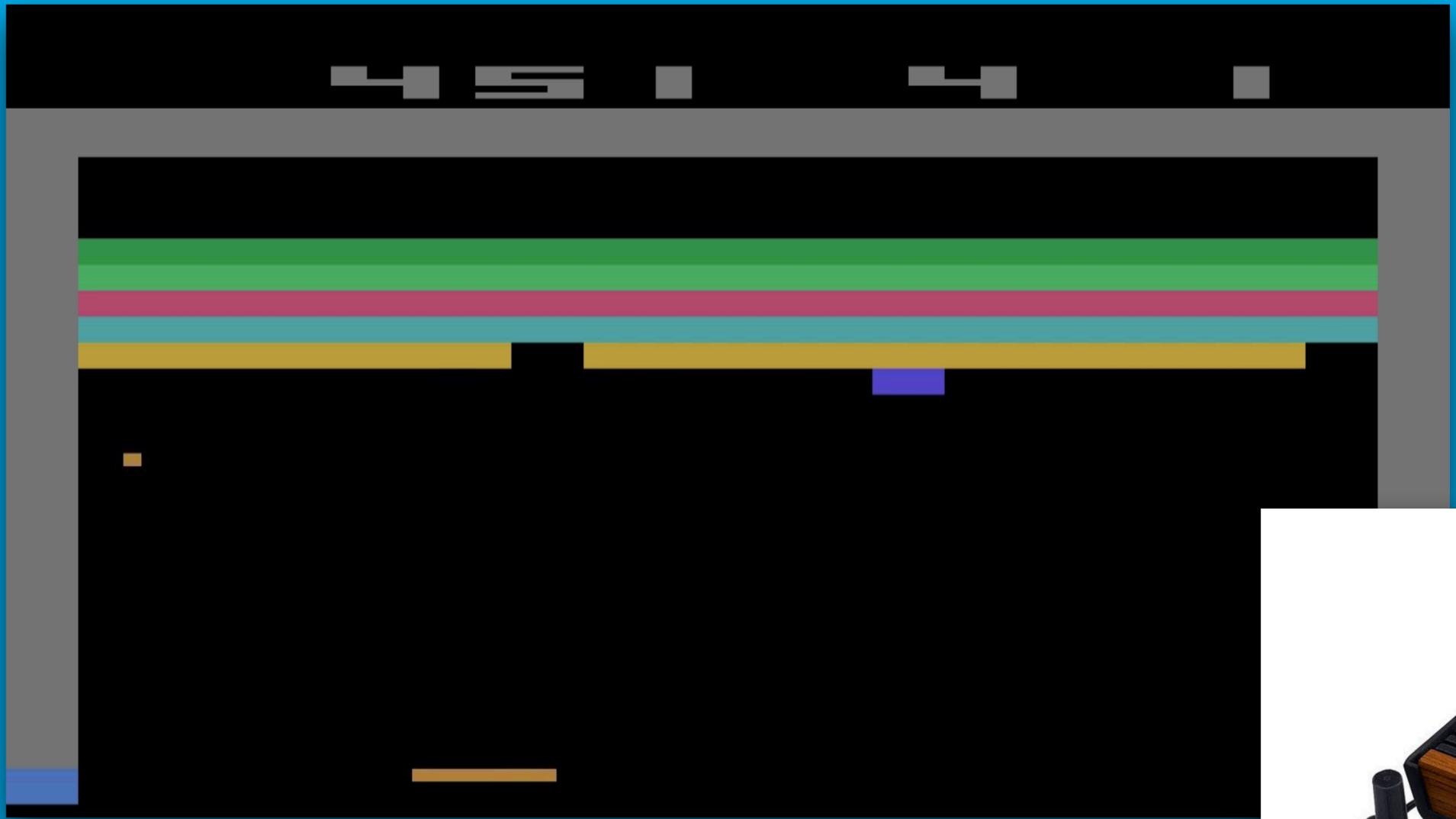
Yves Hilpisch  
[Message](#)

# AI-Powered Algorithmic Trading

- 1. AI Success Stories**
- 2. The Beauty Myth**
- 3. Data-Driven Finance**
- 4. Efficient Markets**
- 5. AI-First Finance**
- 6. Deep Learning**
- 7. The AI Machine**
- 8. Conclusions**

# AI Success Stories

**AI Success Stories  
—Atari Games and  
Reinforcement Learning**



“We present the first deep learning model to successfully learn control policies directly from high-dimensional sensory input using reinforcement learning. The model is a convolutional neural network, trained with a variant of Q-learning, whose input is raw pixels and whose output is a value function estimating future rewards. We apply our method to seven Atari 2600 games from the Arcade Learning Environment, with no adjustment of the architecture or learning algorithm. We find that it outperforms all previous approaches on six of the games and surpasses a human expert on three of them.”

Mnih, V. (2013): “Playing Atari with Deep Reinforcement Learning”. <https://arxiv.org/pdf/1312.5602v1.pdf>

arXiv:1312.5602v1 [cs.LG] 19 Dec 2013

---

## Playing Atari with Deep Reinforcement Learning

---

Volodymyr Mnih Koray Kavukcuoglu David Silver Alex Graves Ioannis Antonoglou  
Daan Wierstra Martin Riedmiller

DeepMind Technologies

{vlad,koray,david,alex.graves,ioannis,daan,martin.riedmiller} @ deepmind.com

### Abstract

We present the first deep learning model to successfully learn control policies directly from high-dimensional sensory input using reinforcement learning. The model is a convolutional neural network, trained with a variant of Q-learning, whose input is raw pixels and whose output is a value function estimating future rewards. We apply our method to seven Atari 2600 games from the Arcade Learning Environment, with no adjustment of the architecture or learning algorithm. We find that it outperforms all previous approaches on six of the games and surpasses a human expert on three of them.

### 1 Introduction

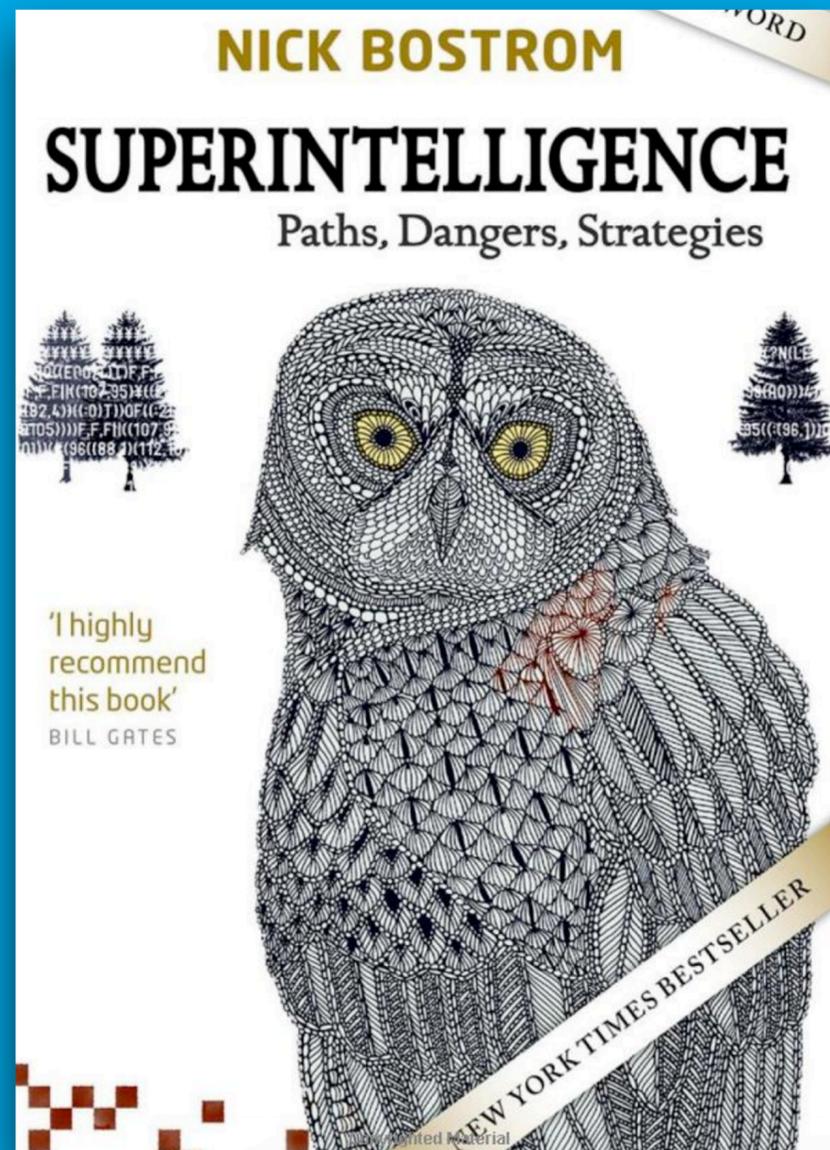
Learning to control agents directly from high-dimensional sensory inputs like vision and speech is one of the long-standing challenges of reinforcement learning (RL). Most successful RL applications that operate on these domains have relied on hand-crafted features combined with linear value functions or policy representations. Clearly, the performance of such systems heavily relies on the quality of the feature representation.

Recent advances in deep learning have made it possible to extract high-level features from raw sensory data, leading to breakthroughs in computer vision [11, 22, 16] and speech recognition [6, 7]. These methods utilise a range of neural network architectures, including convolutional networks, multilayer perceptrons, restricted Boltzmann machines and recurrent neural networks, and have exploited both supervised and unsupervised learning. It seems natural to ask whether similar techniques could also be beneficial for RL with sensory data.

However reinforcement learning presents several challenges from a deep learning perspective. Firstly, most successful deep learning applications to date have required large amounts of hand-labelled training data. RL algorithms, on the other hand, must be able to learn from a scalar reward signal that is frequently sparse, noisy and delayed. The delay between actions and resulting rewards, which can be thousands of timesteps long, seems particularly daunting when compared to the direct association between inputs and targets found in supervised learning. Another issue is that most deep learning algorithms assume the data samples to be independent, while in reinforcement learning one typically encounters sequences of highly correlated states. Furthermore, in RL the data distribution changes as the algorithm learns new behaviours, which can be problematic for deep learning methods that assume a fixed underlying distribution.

This paper demonstrates that a convolutional neural network can overcome these challenges to learn successful control policies from raw video data in complex RL environments. The network is trained with a variant of the Q-learning [26] algorithm, with stochastic gradient descent to update the weights. To alleviate the problems of correlated data and non-stationary distributions, we use

**AI Success Stories  
—Go and AlphaGo**



“Go-playing programs have been improving at a rate of about 1 dan/year in recent years. If this rate of improvement continues, they might beat the human world champion in about a decade.”

*Nick Bostrom (2014): Superintelligence.*

# The story of AlphaGo so far

AlphaGo is the first computer program to defeat a professional human Go player, the first program to defeat a Go world champion, and arguably the strongest Go player in history.

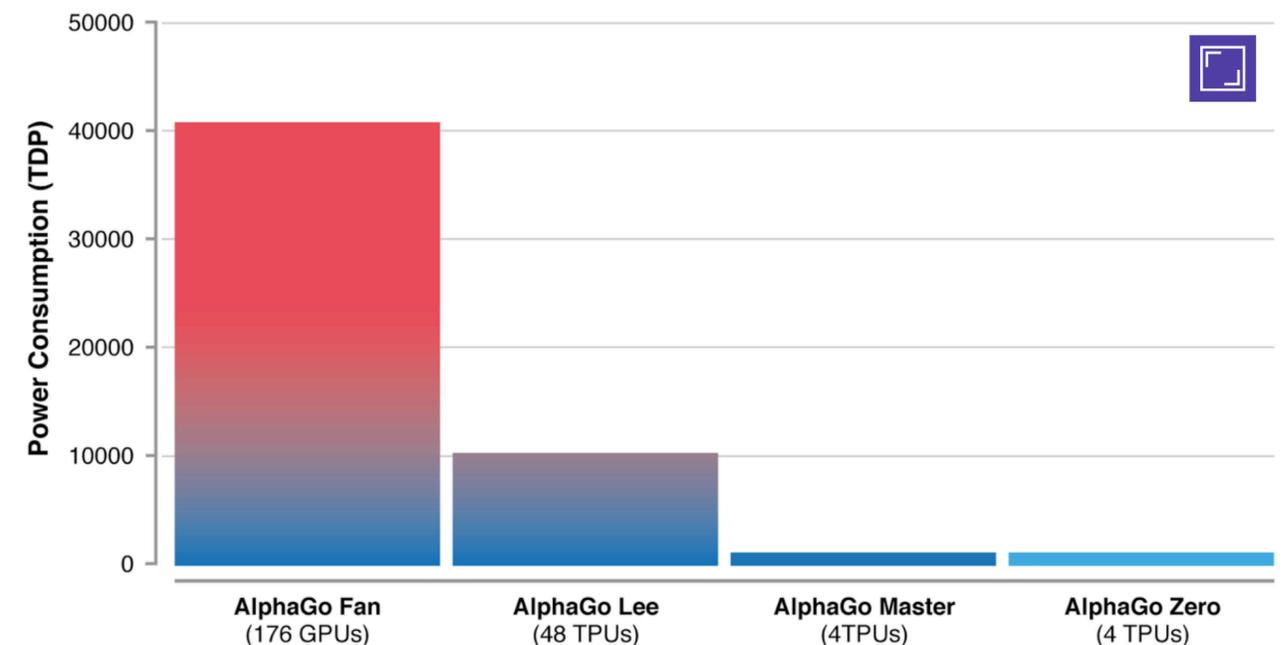
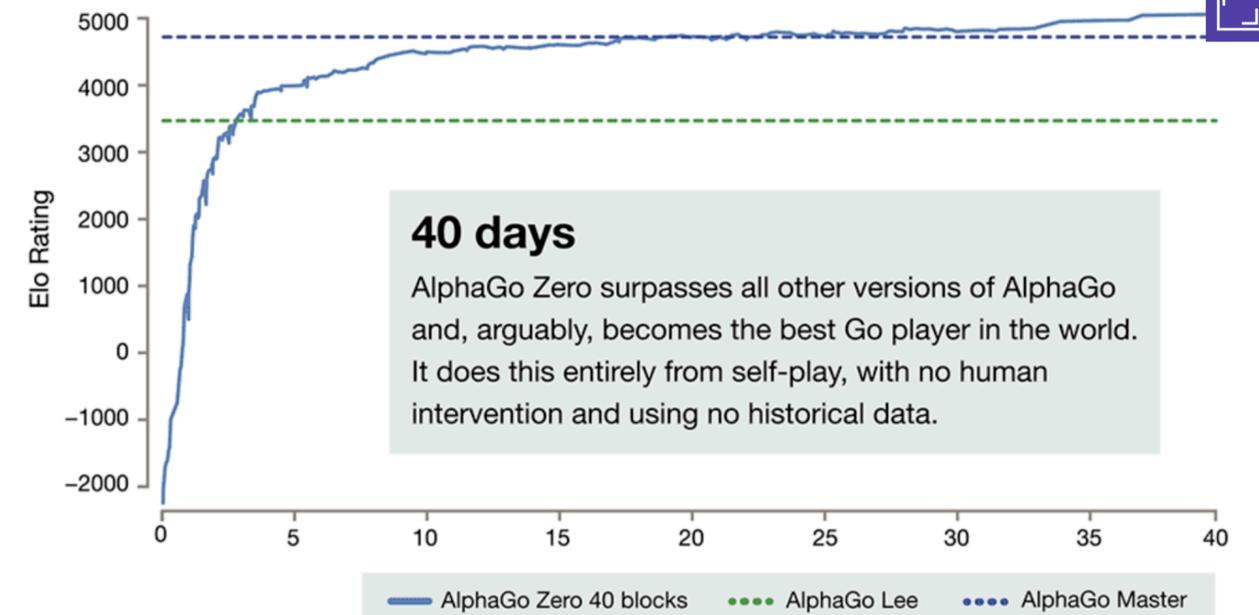
AlphaGo's first formal match was against the reigning 3-times European Champion, Mr Fan Hui, in October 2015. Its 5-0 win was the first ever against a Go professional, and the results were published in full technical detail in the international journal, [Nature](#). AlphaGo then went on to compete against legendary player Mr Lee Sedol, winner of 18 world titles and widely considered to be the greatest player of the past decade.

AlphaGo's 4-1 victory in Seoul, South Korea, in March 2016 was watched by over 200 million people worldwide. It was a landmark achievement that experts agreed was a decade ahead of its time, and earned AlphaGo a 9 dan professional ranking (the highest certification) - the first time a computer Go player had ever received the accolade.

During the games, AlphaGo played a handful of [highly inventive winning moves](#), several of which - including move 37 in game two - were so surprising they overturned hundreds of years of received wisdom, and have since been examined extensively by players of all levels. In the course of winning, AlphaGo somehow taught the world completely new knowledge about perhaps the most studied and contemplated game in history.

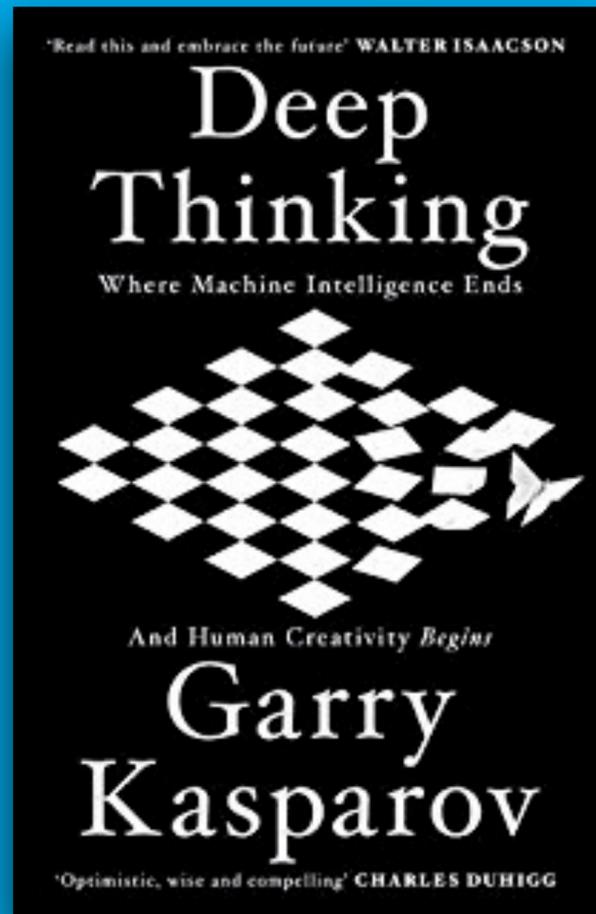
contemplated game in history.

taught the world completely new knowledge about perhaps the most studied and extensively played games of all levels. In the course of winning, AlphaGo somehow



AlphaGo has become progressively more efficient thanks to hardware gains and more recently algorithmic advances

**AI Success Stories**  
**—Chess, Deep Blue & AlphaZero**



“It was a pleasant day in Hamburg in June 6, 1985, ... Each of my opponents, all thirty-two of them, was a computer. ... it didn't come as much of a surprise, ..., when I achieved a perfect 32—0 score.”

“Twelve years later I was in New York City fighting for my chess life. Against just one machine, a \$10 million IBM supercomputer nicknamed ‘Deep Blue’.”

“Jump forward another 20 years to today, to 2017, and you can download any number of free chess apps for your phone that rival any human Grandmaster.”

A dark purple rectangular banner with a background image of a chessboard and chess pieces. The text is white and centered.

## AlphaZero: Shedding new light on the grand games of chess, shogi and Go

“Traditional chess engines — including the world computer chess champion Stockfish and IBM’s ground-breaking Deep Blue — rely on **thousands of rules and heuristics handcrafted by strong human players** that try to account for every eventuality in a game. ...

AlphaZero takes a totally different approach, replacing these hand-crafted rules with a **deep neural network** and **general purpose algorithms** that know nothing about the game beyond the basic rules.”

“The amount of **training** the network needs depends on the style and complexity of the game, taking **approximately 9 hours for chess**, 12 hours for shogi, and 13 days for Go.”

“In Chess, for example, it searches **only 60 thousand positions** per second in chess, compared to roughly 60 million for Stockfish.”

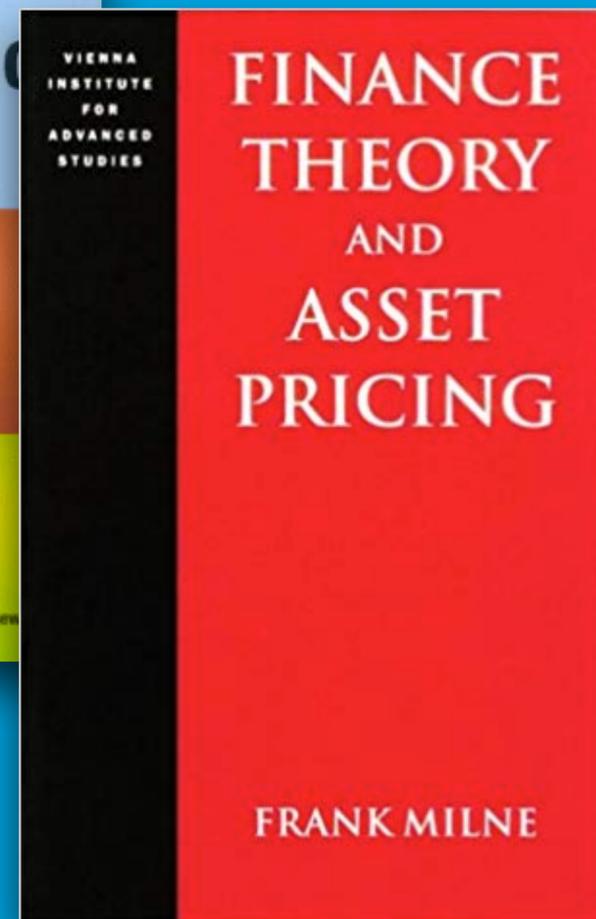
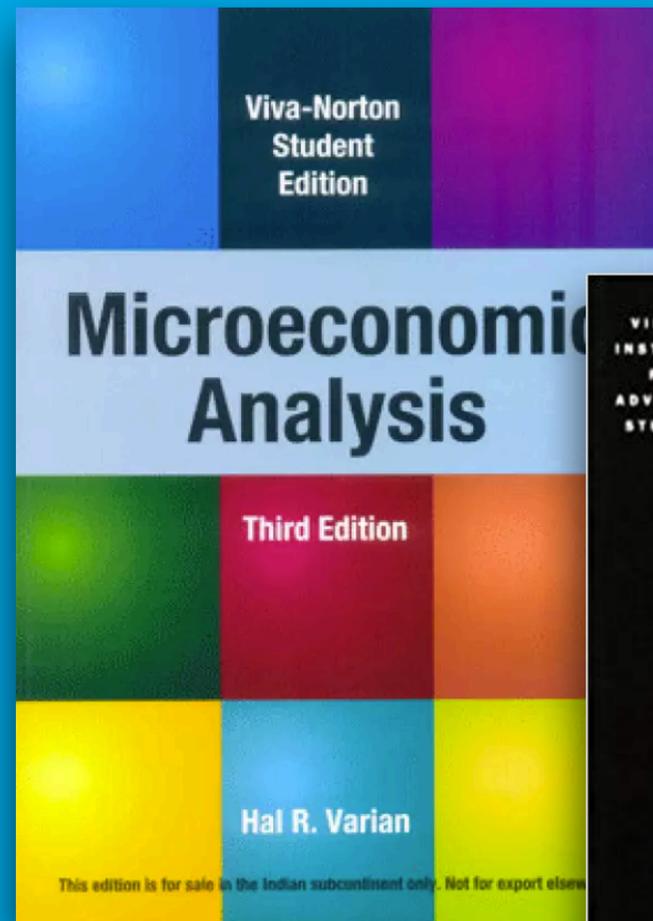
Source: <http://deepmind.com>

# The Beauty Myth

## Fundamental physics is frustrating physicists

GUTs are among several long-established theories that remain stubbornly unsupported by the big, costly experiments testing them. ...

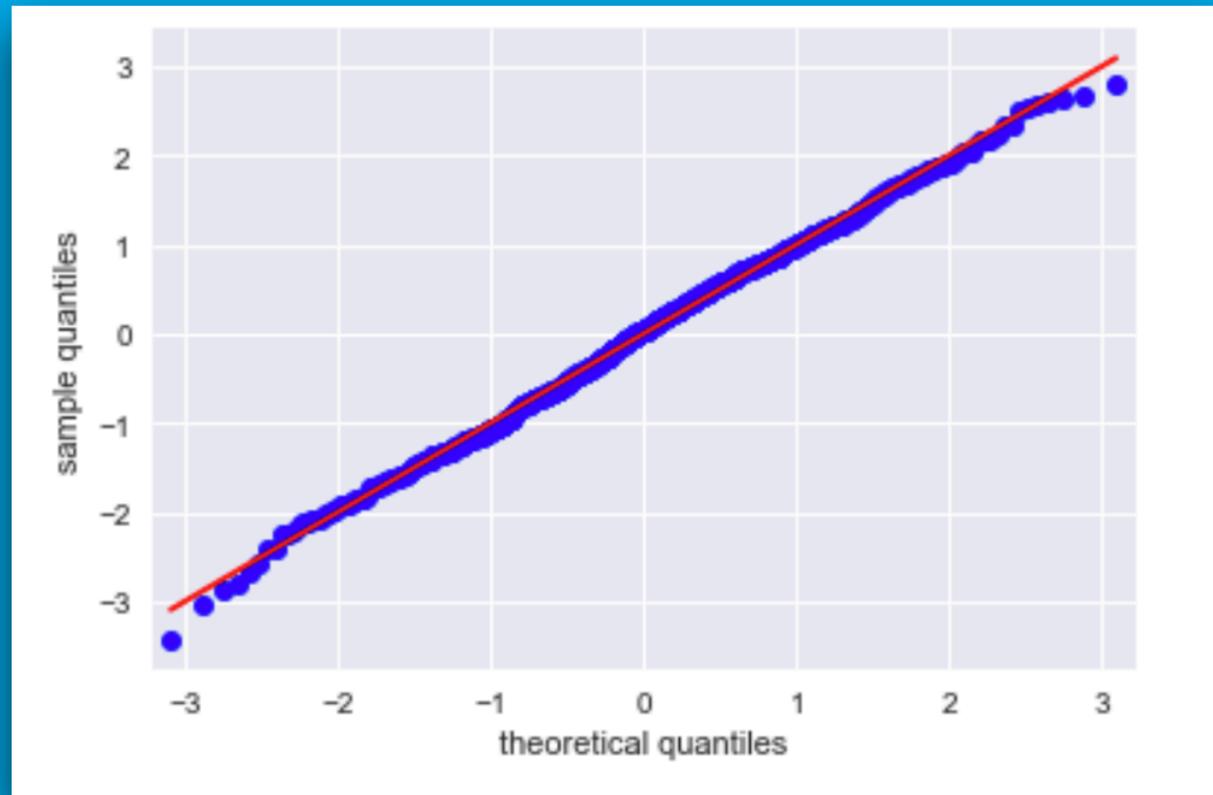
Despite the dearth of data, the answers that all these theories offer to some of the most vexing questions in physics are so elegant that they populate postgraduate textbooks. As Peter Woit of Columbia University observes, “Over time, these ideas became institutionalised. People stopped thinking of them as speculative.” That is understandable, for they appear to have great explanatory power.



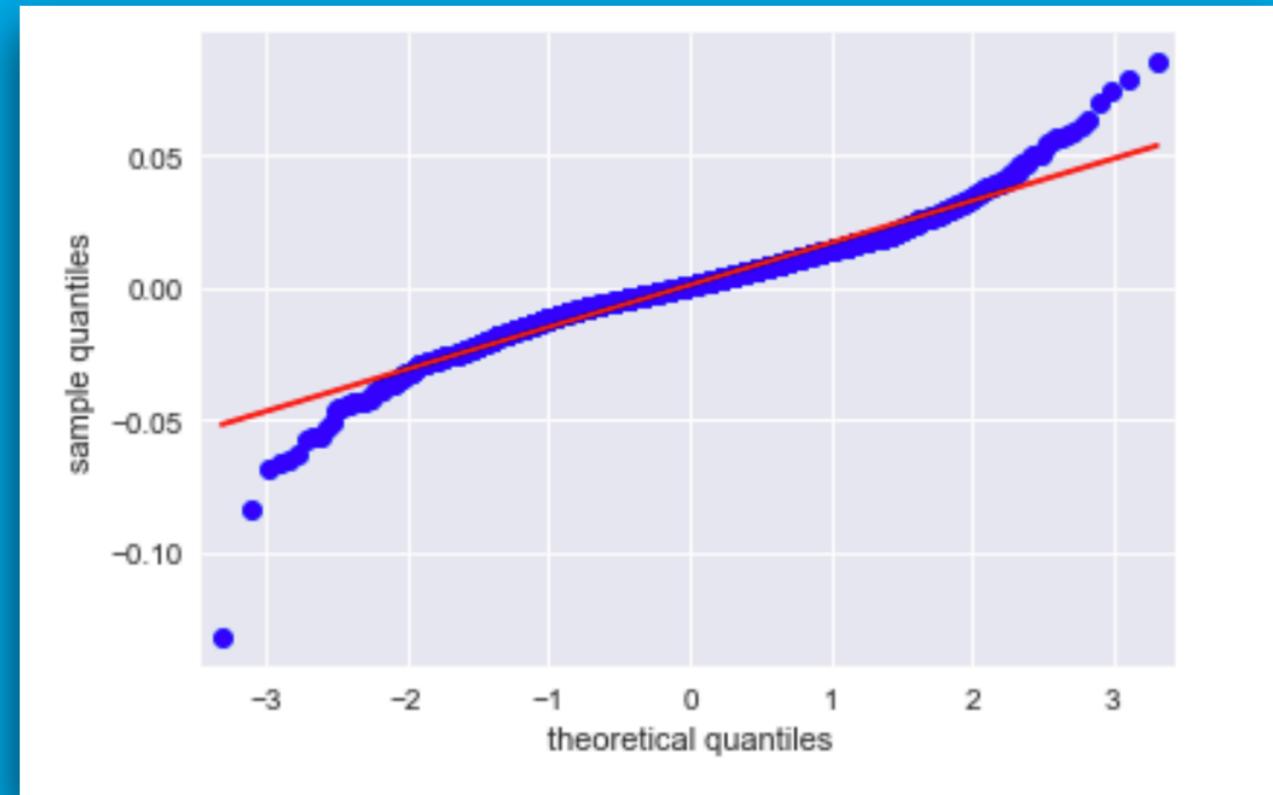
## Cornerstones of Economics

- A. Arbitrage Pricing
- B. Expected Utility
- C. Equilibrium
- D. Normal Distributions
- E. Linear Relationships
- F. Efficient Markets

# Theory



# Reality



CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK\*

WILLIAM F. SHARPE†

I. INTRODUCTION

ONE OF THE PROBLEMS which has plagued those attempting to predict the behavior of capital markets is the absence of a body of positive micro-economic theory dealing with conditions of risk. Although many useful insights can be obtained from the traditional models of investment under conditions of certainty, the pervasive influence of risk in financial transactions has forced those working in this area to adopt models of price behavior which are little more than assertions. A typical classroom explanation of the determination of capital asset prices, for example, usually begins with a careful and relatively rigorous description of the process through which individual preferences and physical relationships interact to determine an equilibrium pure interest rate. This is generally followed by the assertion that somehow a market risk-premium is also determined, with the prices of assets adjusting accordingly to account for differences in their risk.

A useful representation of the view of the capital market implied in such discussions is illustrated in Figure 1. In equilibrium, capital asset prices have adjusted so that the investor, if he follows rational procedures (primarily diversification), is able to attain any desired point along a *capital market line*.<sup>1</sup> He may obtain a higher expected rate of return on his holdings only by incurring additional risk. In effect, the market presents him with two prices: the *price of time*, or the pure interest rate (shown by the intersection of the line with the horizontal axis) and the *price of risk*, the additional expected return per unit of risk borne (the reciprocal of the slope of the line).

\* A great many people provided comments on early versions of this paper which led to major improvements in the exposition. In addition to the referees, who were most helpful, the author wishes to express his appreciation to Dr. Harry Markowitz of the RAND Corporation, Professor Jack Hirshleifer of the University of California at Los Angeles, and to Professors Yoram Barzel, George Brabb, Bruce Johnson, Walter Oi and R. Haney Scott of the University of Washington.

† Associate Professor of Operations Research, University of Washington.

1. Although some discussions are also consistent with a non-linear (but monotonic) curve.

$$\mu_i = r + \beta_i(\mu_M - r)$$

“Market Risk”

“Idiosyncratic Risk”

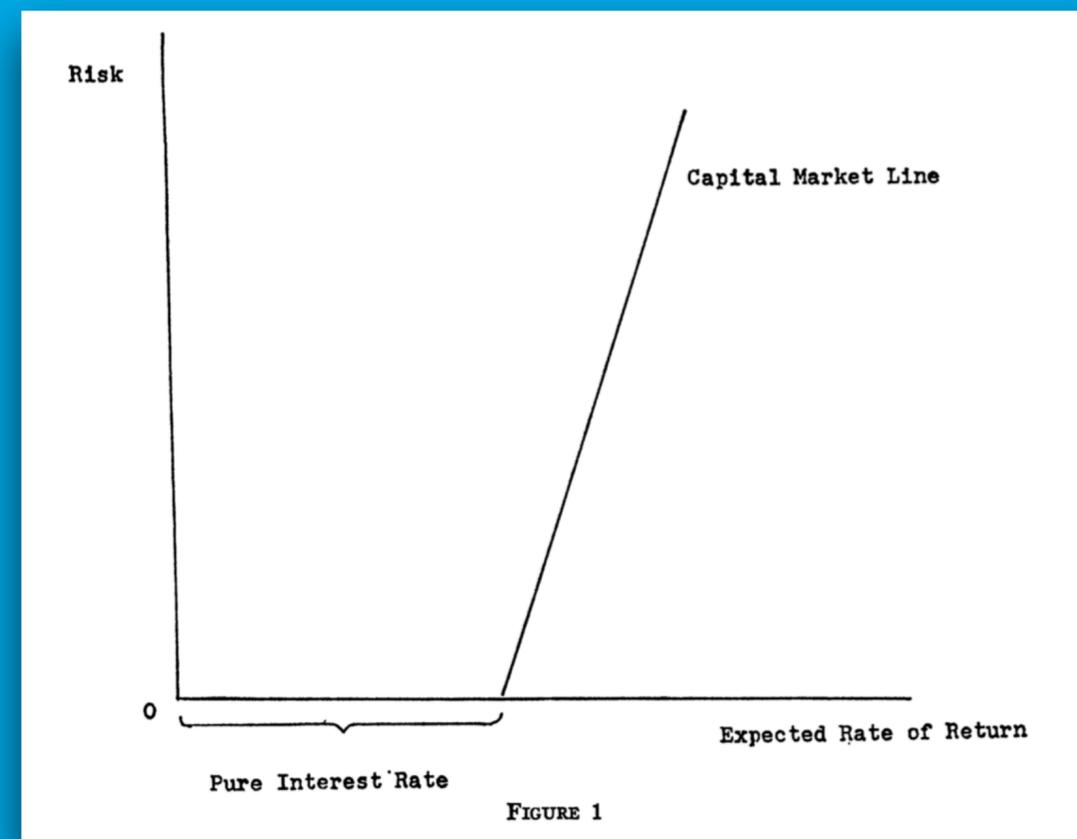


FIGURE 1

# Data-Driven Finance



**BUSINESS SUMMARY** >

Apple Inc. designs, manufactures and markets mobile communication and media devices, personal computers and portable digital music players. The Company sells a range of related software, services, accessories, networking solutions, and third-party digital content and applications. The Company's segments include the Americas, Europe, Greater China, Japan and Rest of Asia Pacific. The Americas segment includes both North and South America. The Europe segment includes European countries, India, the Middle East and Africa. The Greater China segment includes China, Hong Kong and Taiwan. The Rest of Asia Pacific segment includes Australia and the Asian countries not included in the Company's other operating segments. Its products and services include iPhone, iPad, Mac, iPod, Apple Watch, Apple TV, a portfolio of consumer and professional software applications, iPhone OS (iOS), OS X and watchOS operating systems, iCloud, Apple Pay and a range of accessory, service and support offerings.

**NEWS** >

28-Dec-2016

10:24:36	Apple dominerade julhandeln mätt i antalet aktiverade enheter	FNW
10:15:18	UPDATE 3-S.Korea fines Qualcomm \$854 mln for violating competition laws	RTRS
09:42:52	Corea del Sur multa a Qualcomm con 854 mlns dlr por violar leyes de competencia	RTRS
06:00:10	RPT-Wall Street cale une fois de plus au seuil des 20.000 points	RTRS
03:30:18	Aumento del gasto de último minuto impulsa a temporada de ventas de fin de año ...	RTRS
01:50:14	Last-minute spending surge lifts U.S. holiday shopping season	RTRS

27-Dec-2016

23:33:16	Reuters Insider - Tech stocks could take the Dow to 20k	CNBC
23:32:28	Reuters Insider - History suggests Dow could hit 20k by Friday: Technician	CNBC
22:55:29	LEAD 2-Wall Street cale une fois de plus au seuil des 20.000 points	RTRS
22:09:39	Apple, Cisco Lead DJIA Higher Tuesday	WALLST

**EVENTS** > Upcoming Past

24-Jan-2017 > 30-Jan-2017

NTS	Q1 2017 Apple Inc Earnings Release	📅
-----	------------------------------------	---

24-Feb-2017 > 28-Feb-2017

**PRICE PERFORMANCE** >

Open	
Prev. Close	
Bid / Ask	
VWAP	--
Turnover	
Volume	
Short Interest	0.90%
YTD	
Beta (5Y Monthly)	1.29
Mkt Cap	USD 625.27B
PE (LTM)	14.12
Div Yield	1.94%
DR	BRL <b>AAPL34.SA</b> (1:0.1)
DR Type	--
DR Bank	--



**AAPL.O 115.190000**

Price USD

117.260000

115.00

110.00

105.00

100.00

95.00

90.00

85.00

Dec-31 Mar-31 Jun-30 Sep-14-Dec-2016

Today 5D 3M 6M 1Y 5Y No Benchmark ▼

52Wk: 89.47 12-May 118.69 11-Oct

📅 Next Earn Report: 24-Jan-2017

Free Float	5.32B	Asset Type	Ordinary Share	5 yr CDS	26.980 bps
Outstanding	5.33B	Share Class	--	Δ Today	-0.07%
IPO Date	12-Dec-1980	Lot Size		Δ 1 Week	-0.074
First Trade Da...	12-Dec-1980	Voting Rights	1		

**FUNDAMENTALS** >

	AAPL (Sep-2016)	Growth	Industry
Gross Margin	38.02%	(4.71%) 4Q	38.91%
Operating Margin	25.10%	(11.59%) 4Q	5.75%

```
In [28]: data = ek.get_timeseries('AAPL.O',
                                fields='*',
                                start_date='2019-06-28 16:00:00',
                                end_date='2019-06-28 17:00:00',
                                interval='tick')
```

```
In [29]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 20530 entries, 2019-06-28 16:00:00.132000 to 2019-06-28 16:59:58.984000
Data columns (total 2 columns):
VALUE      20499 non-null float64
VOLUME     20530 non-null float64
dtypes: float64(2)
memory usage: 481.2 KB
```

```
In [30]: data.head()
```

```
Out[30]:
```

	AAPL.O	VALUE	VOLUME
	Date		
	2019-06-28 16:00:00.132	198.64	100.0
	2019-06-28 16:00:00.132	198.64	100.0
	2019-06-28 16:00:00.132	198.64	10.0
	2019-06-28 16:00:00.133	198.64	400.0
	2019-06-28 16:00:00.133	198.64	1300.0

```
In [31]: news = ek.get_news_headlines('R:TSLA.O PRODUCTION',
                                     date_from='2019-07-01',
                                     date_to='2019-07-19',
                                     count=5
                                     )
```

```
In [32]: news
```

```
Out[32]:
```

	versionCreated		text	storyId	sourceCode
	2019-07-18 07:37:32.841	2019-07-18 07:37:32.841	Tesla takes production global	urn:newsml:reuters.com:20190718:nNRA98d8a7:1	NS:GLOBML
	2019-07-16 15:55:30.319	2019-07-16 15:55:30.319	Reuters Insider - Tesla workers say they use s...	urn:newsml:reuters.com:20190716:nRTV7qNh5s:1	NS:CNBC
	2019-07-11 14:20:18.249	2019-07-11 14:20:18.249	Auto Stock Roundup: TSLA to Raise Production, ...	urn:newsml:reuters.com:20190711:nNRA96h0jy:1	NS:ZACKSC
	2019-07-11 13:15:45.895	2019-07-11 13:15:45.895	Tesla (TSLA) Prepares to Increase Production i...	urn:newsml:reuters.com:20190711:nNRA96ghdy:1	NS:ZACKSC
	2019-07-11 08:36:17.569	2019-07-11 08:36:17.569	Tesla signals it's in hiring mode The company ...	urn:newsml:reuters.com:20190711:nNRA96d7au:1	NS:LATIME

```
In [33]: storyId = news['storyId'][0]
```

```
In [34]: from IPython.display import HTML
```

```
In [35]: HTML(ek.get_news_story(storyId))
```

```
Out[35]: A new Tesla factory is seen under construction in Shanghai on Tuesday. The electric-car manufacturer is building its first overseas car factory to accelerate its annual production
```



## EXPERT OPINION

Contact Editor: **Brian Brannon**, bbrannon@computer.org

# The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

Eugene Wigner's article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences"<sup>1</sup> examines why so much of physics can be neatly explained with simple mathematical formulas

such as  $f = ma$  or  $e = mc^2$ . Meanwhile, sciences that involve human beings rather than elementary particles have proven more resistant to elegant mathematics. Economists suffer from physics envy over their inability to neatly model human behavior. An informal, incomplete grammar of the English language runs over 1,700 pages.<sup>2</sup> Perhaps when it comes to natural language processing and related fields, we're doomed to complex theories that will never have the elegance of physics equations. But if that's so, we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.

One of us, as an undergraduate at Brown University, remembers the excitement of having access to the Brown Corpus, containing one million English words.<sup>3</sup> Since then, our field has seen several notable corpora that are about 100 times larger, and in 2006, Google released a trillion-word corpus with frequency counts for all sequences up to five words long.<sup>4</sup> In some ways this corpus is a step backwards from the Brown Corpus: it's taken from unfiltered Web pages and thus contains incomplete sentences, spelling errors, grammatical errors, and all sorts of other errors. It's not annotated with carefully hand-corrected part-of-speech tags. But the fact that it's a million times larger than the Brown Corpus outweighs these drawbacks. A trillion-word corpus—along with other Web-derived corpora of millions, billions, or trillions of links, videos, images, tables, and user interactions—captures even very rare aspects of human

behavior. So, this corpus could serve as the basis of a complete model for certain tasks—if only we knew how to extract the model from the data.

### Learning from Text at Web Scale

The biggest successes in natural-language-related machine learning have been statistical speech recognition and statistical machine translation. The reason for these successes is not that these tasks are easier than other tasks; they are in fact much harder than tasks such as document classification that extract just a few bits of information from each document. The reason is that translation is a natural task routinely done every day for a real human need (think of the operations of the European Union or of news agencies). The same is true of speech transcription (think of closed-caption broadcasts). In other words, a large training set of the input-output behavior that we seek to automate is available to us *in the wild*. In contrast, traditional natural language processing problems such as document classification, part-of-speech tagging, named-entity recognition, or parsing are not routine tasks, so they have no large corpus available in the wild. Instead, a corpus for these tasks requires skilled human annotation. Such annotation is not only slow and expensive to acquire but also difficult for experts to agree on, being bedeviled by many of the difficulties we discuss later in relation to the Semantic Web. The first lesson of Web-scale learning is to use available large-scale data rather than hoping for annotated data that isn't available. For instance, we find that useful semantic relationships can be automatically learned from the statistics of search queries and the corresponding results<sup>5</sup> or from the accumulated evidence of Web-based text patterns and formatted tables,<sup>6</sup> in both cases without needing any manually annotated data.

Eugene Wigner's article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" examines why so much of physics can be neatly explained with simple mathematical formulas such as  $f = ma$  or  $e = mc^2$ . Meanwhile, sciences that involve human beings rather than elementary particles have proven more resistant to elegant mathematics. Economists suffer from physics envy over their inability to neatly [and successfully] model human behavior. An informal, incomplete grammar of the English language runs over 1,700 pages. Perhaps when it comes to natural language processing and related fields, we're doomed to complex theories that will never have the elegance of physics equations. But if that's so, we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.

# How investment analysts became data miners

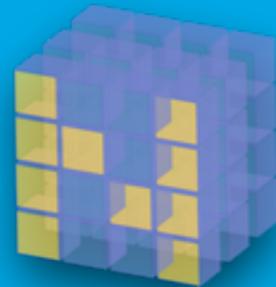
Banks battle for audiences with new information sets, 'charticles' and podcasts



© AP



```
IP[y]:  
IPython
```



NumPy

matplotlib



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



xlwings



# Efficient Markets

## Random Walks in Stock Market Prices

Eugene F. Fama

For many years economists, statisticians, and teachers of finance have been interested in developing and testing models of stock price behavior. One important model that has evolved from this research is the theory of random walks. This theory casts serious doubt on many other methods for describing and predicting stock price behavior—methods that have considerable popularity outside the academic world. For example, we shall see later that if the random walk theory is an accurate description of reality, then the various “technical” or “chartist” procedures for predicting stock prices are completely without value.

In general the theory of random walks raises challenging questions for anyone who has more than a passing interest in understanding the behavior of stock prices. Unfortunately, however, most discussions of the theory have appeared in technical academic journals and in a form which the non-mathematician would usually find incomprehensible. This article describes, briefly and simply, the theory of random walks and some of the important issues it raises concerning the work of market analysts. To preserve brevity some aspects of the theory and its implications are omitted. More complete (and also more technical) discussions of the theory of random walks are available elsewhere; hopefully the introduction provided here will encourage the reader to examine one of the more rigorous and lengthy works listed at the end of this article.

### COMMON TECHNIQUES FOR PREDICTING STOCK MARKET PRICES

In order to put the theory of random walks into perspective we first discuss, in brief and general terms, the two approaches to predicting stock prices that are commonly espoused by market professionals. These are (1) “chartist” or “technical” theories and (2) the theory of fundamental or intrinsic value analysis.

The basic assumption of all the chartist or technical theories is that history tends to repeat

itself, i.e., past patterns of price behavior in individual securities will tend to recur in the future. Thus the way to predict stock prices (and, of course, increase one’s potential gains) is to develop a familiarity with past patterns of price behavior in order to recognize situations of likely recurrence.

Essentially, then, chartist techniques attempt to use knowledge of the past behavior of a price series to predict the probable future behavior of the series. A statistician would characterize such techniques as assuming that successive price changes in individual securities are dependent. That is, the various chartist theories assume that the *sequence* of price changes prior to any given day is important in predicting the price change for that day.<sup>1</sup>

The techniques of the chartist have always been surrounded by a certain degree of mysticism, however, and as a result most market professionals have found them suspect. Thus it is probably safe to say that the pure chartist is relatively rare among stock market analysts. Rather the typical analyst adheres to a technique known as fundamental analysis or the intrinsic value method. The assumption of the fundamental analysis approach is that at any point in time an individual security has an intrinsic value (or in the terms of the economist, an equilibrium price) which depends on the earning potential of the security. The earning potential of the security depends in turn on such fundamental factors as quality of management, outlook for the industry and the economy, etc.

Through a careful study of these fundamental factors the analyst should, in principle, be able to determine whether the actual price of a security is above or below its intrinsic value. If actual prices tend to move toward intrinsic values, then attempting to determine the intrinsic value of a security is equivalent to making a prediction of its future price; and this is the essence of the predictive procedure implicit in fundamental analysis.

### THE THEORY OF RANDOM WALKS

Chartist theories and the theory of fundamental analysis are really the province of the market

## Eugene F. Fama (1965):

“For many years, economists, statisticians, and teachers of finance have been interested in developing and testing models of stock price behavior. One important model that has evolved from this research is the theory of random walks. This theory casts serious doubt on many other methods for describing and predicting stock price behavior—methods that have considerable popularity outside the academic world. For example, we shall see later that, if the random-walk theory is an accurate description of reality, then the various “technical” or “chartist” procedures for predicting stock prices are completely without value.”—Eugene F. Fama (1965): “Random Walks in Stock Market Prices”

Reprinted from *Financial Analysts Journal* (September/October 1965):55-59.

## **Michael Jensen (1978): “Some Anomalous Evidence Regarding Market Efficiency”:**

“A market is efficient with respect to an information set  $S$  if it is impossible to make economic profits by trading on the basis of information set  $S$ .”

If a stock price follows a (simple) random walk (no drift & normally distributed returns), then it rises and falls with the same probability of 50% (“toss of a coin”).

**In such a case, the best predictor of tomorrow’s stock price — in a least-squares sense — is today’s stock price.**

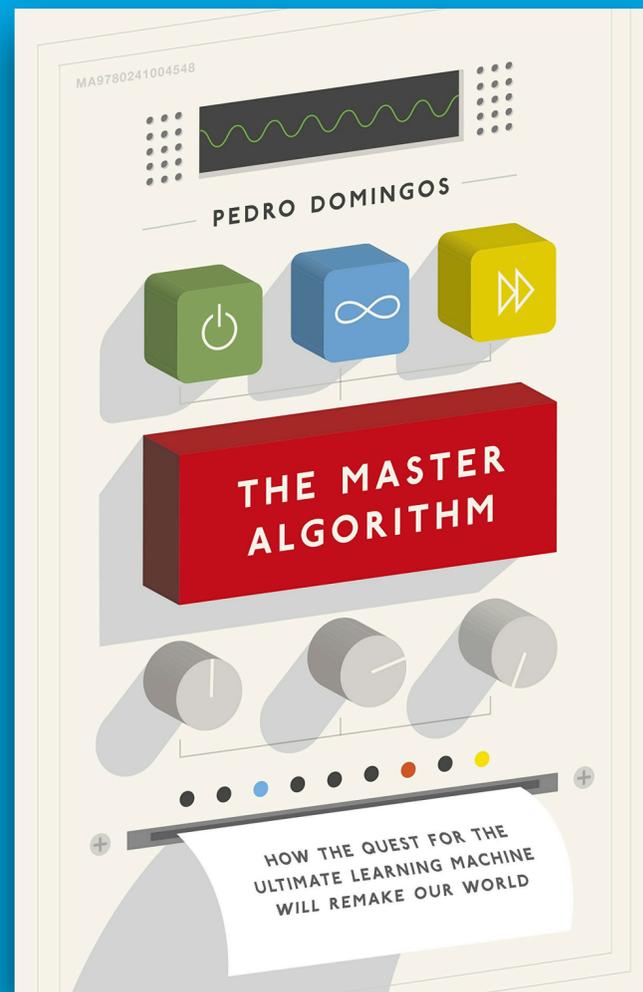
**AI-First Finance**

# scientific method

*noun*

a method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses.

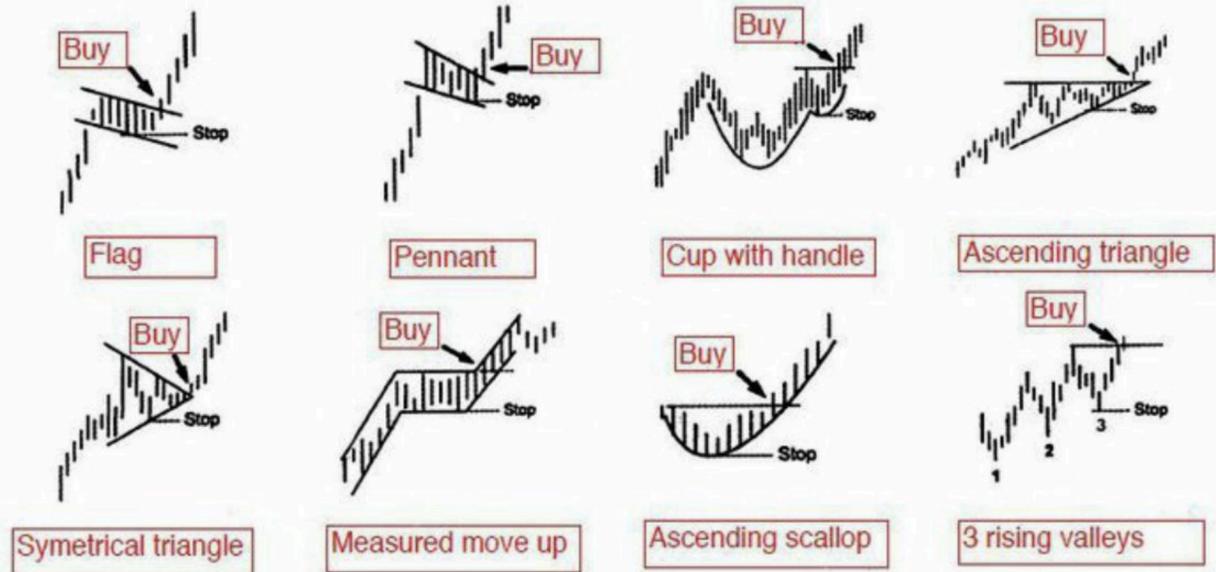
"criticism is the backbone of **the scientific method**"



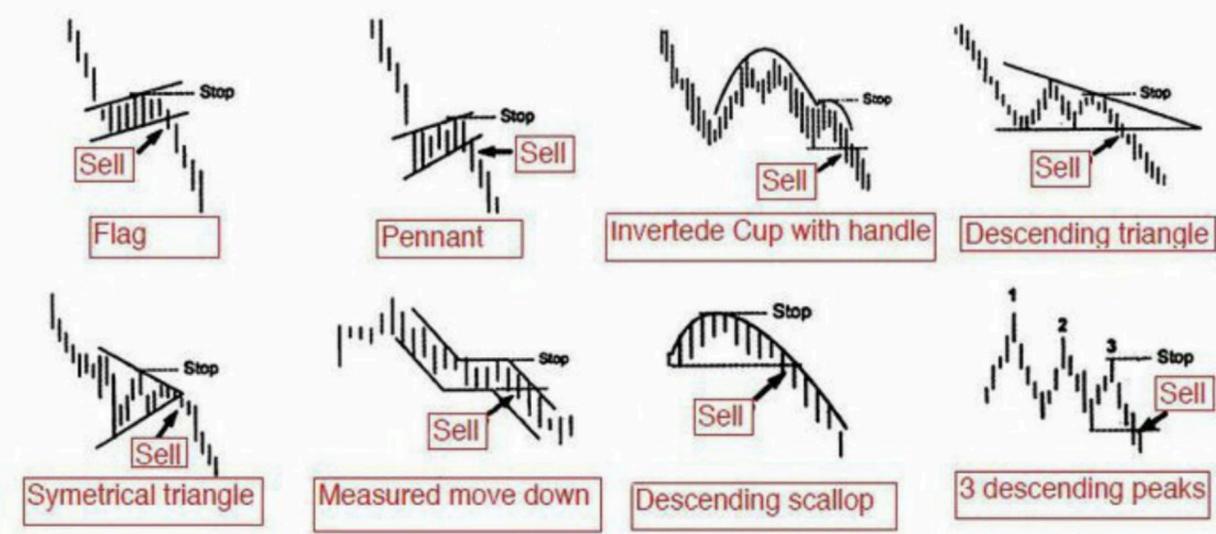
“The grand aim of science is to cover the greatest number of experimental facts by logical deduction from the smallest number of hypotheses or axioms.”  
— Albert Einstein

“Machine learning is the scientific method on steroids. It follows the same process of generating, testing, and discarding or refining hypotheses. But while a scientist may spend his or her whole life coming up with and testing a few hundred hypotheses, a machine-learning system can do the same in a second. Machine learning automates discovery. It’s no surprise, then that it’s revolutionizing science as much as it’s revolutionizing business.”

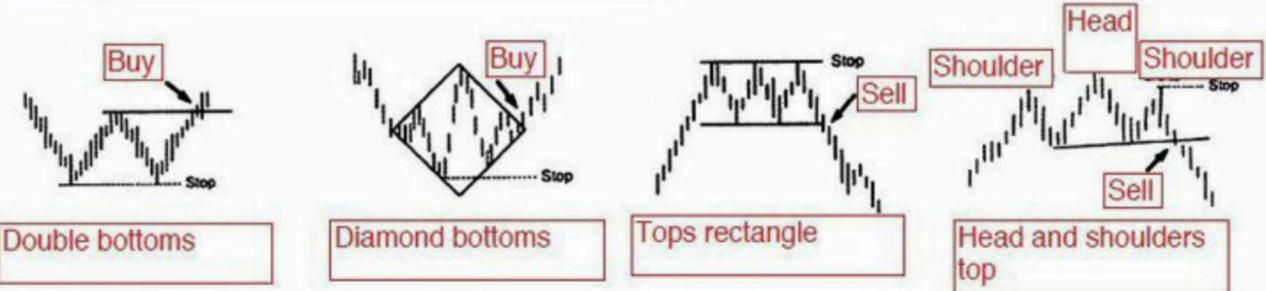
Bullish patterns (going up)



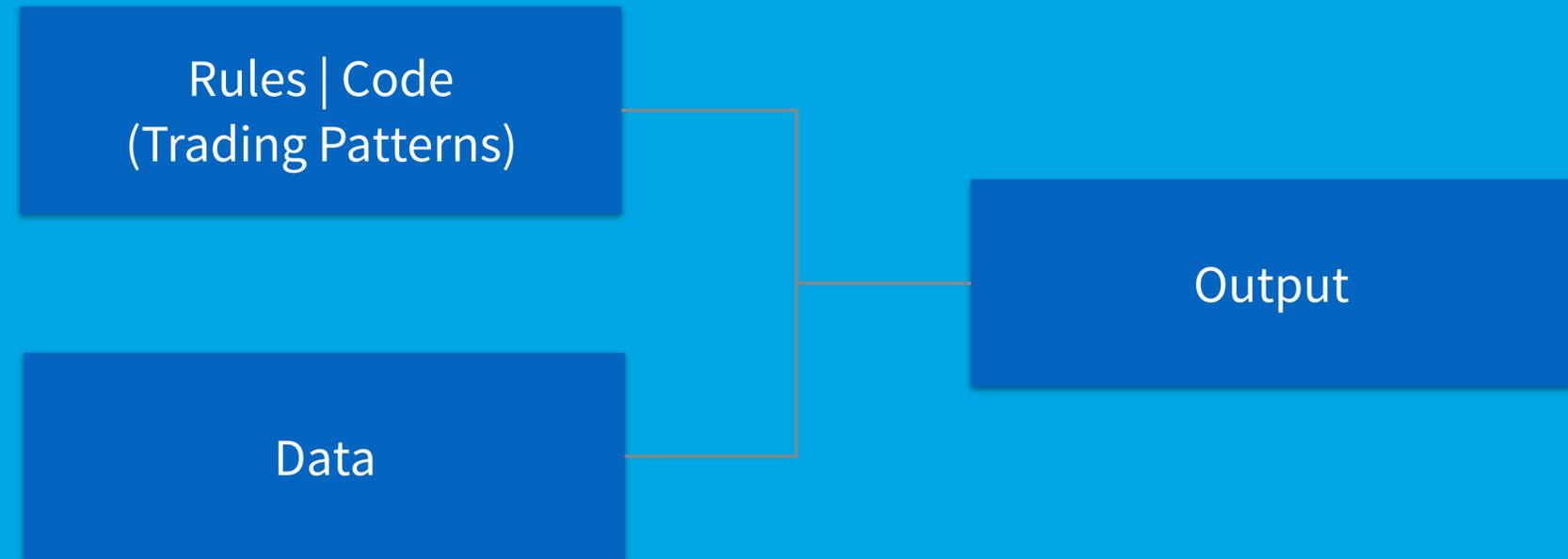
Bearish patterns (going down)



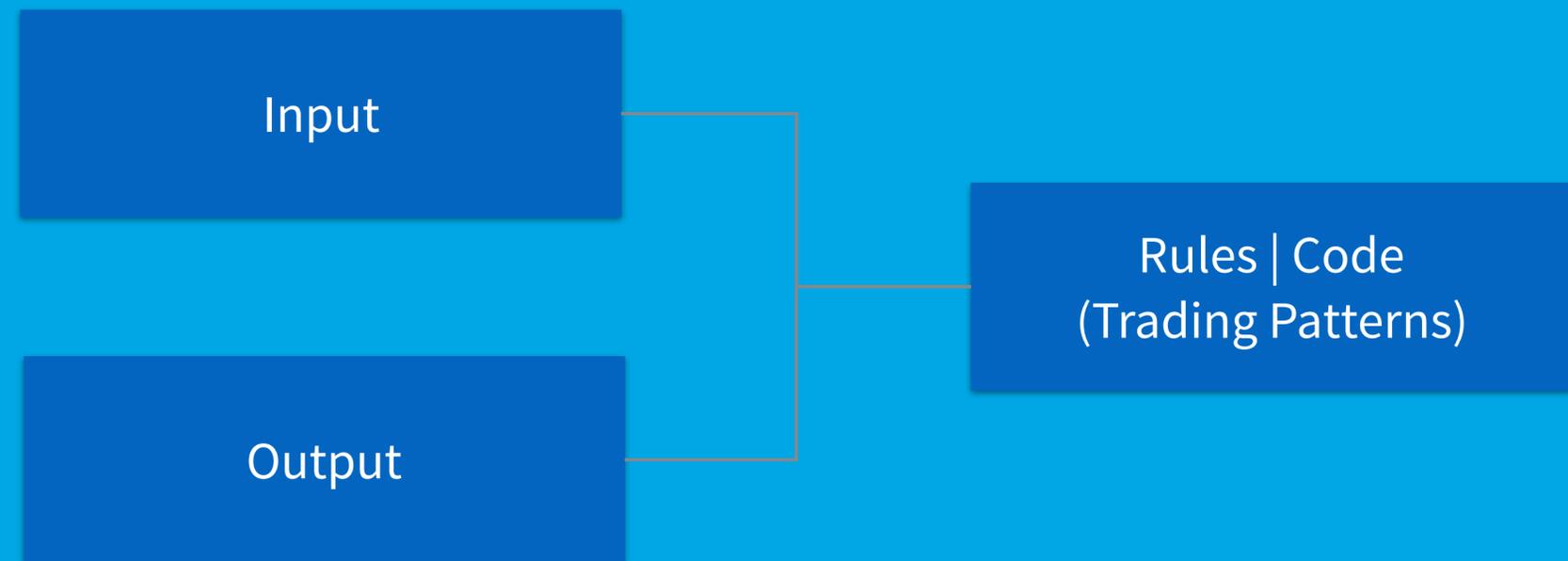
Reversal patterns



## Programming.



## Machine Learning.



# Financial Markets

“normative economics = assumptions, axioms, etc.”

$x$



$y$

“non-linear, complex, changing”

# Finance History



(too) “simple and elegant theories”

$f(\cdot)$

$f(x) \neq y$

“brain-driven & beauty myth”

# AI in Finance = finance

“positive economics = data, relationships, etc.”

$x$

“general, parametrizable, trainable algorithms”

$m(\cdot, a, b)$

“might show good performance, but black box”

$m(x, a^*, b^*) \approx y$

“data-driven & AI-first”

MARCOS LOPEZ DE PRADO

ADVANCES  
*in*  
FINANCIAL  
MACHINE  
LEARNING

WILEY

“The essential tool of econometrics is multivariate linear regression, an 18th-century technology that was already mastered by Gauss before 1794 ... It is hard to believe that something as complex as 21st-century finance could be grasped by something as simple as inverting a covariance matrix.”

“... what if economists finally started to consider non-linear functions?”

“An ML algorithm can spot patterns in a 100-dimensional world as easily as in our familiar 3-dimensional one.”

“Econometrics might be good enough to succeed in financial academia (for now), but succeeding in practice requires ML.”

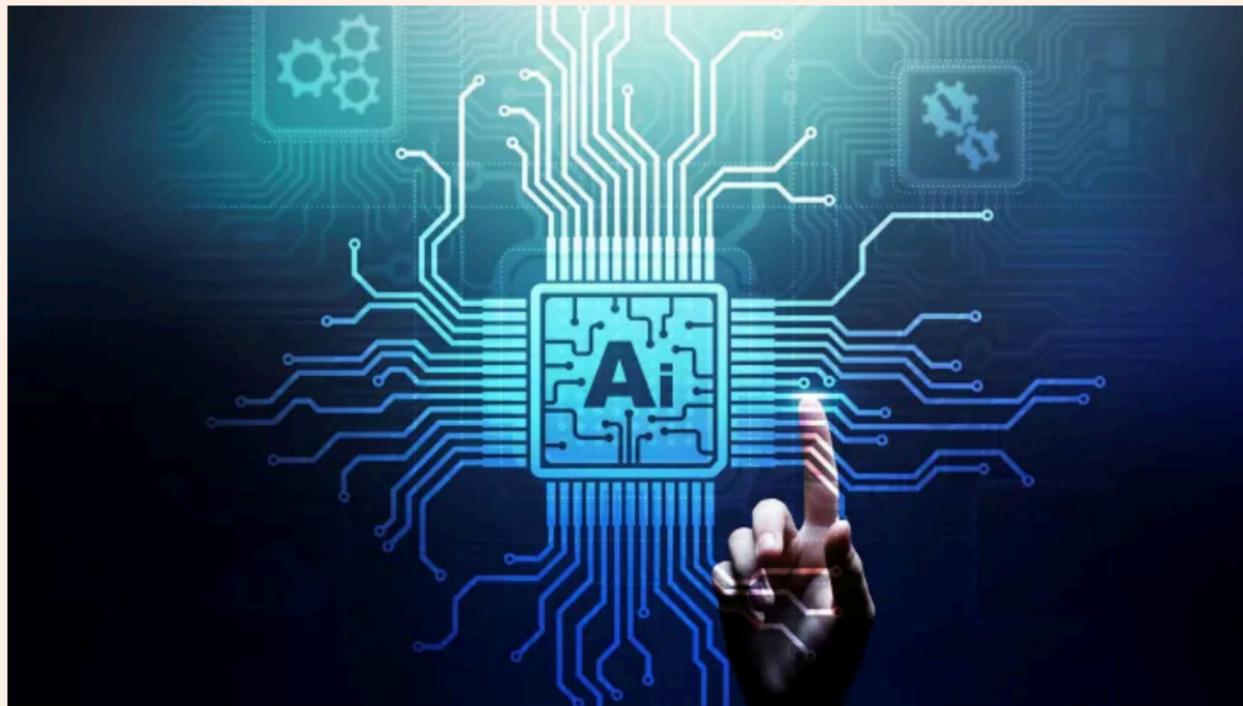
*Marcos López de Prado (2018)*

# Will bots replace humans in active equity investment?

Machine learning will implement quant analysis to find the right stocks

ROBERT POZEN

[+ Add to myFT](#)



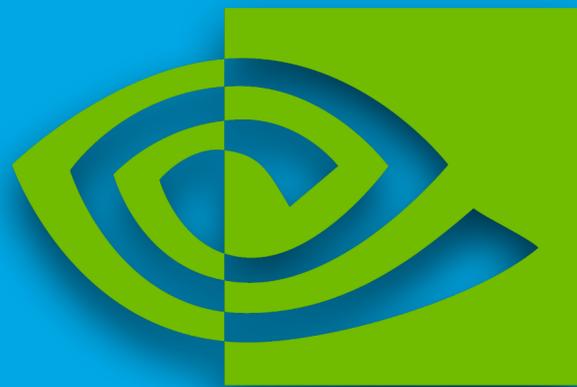
ML can identify outperforming equities based on patterns that would have not been selected by humans © Dreamstime

# How quants and QE shook the cult of the stockpicker

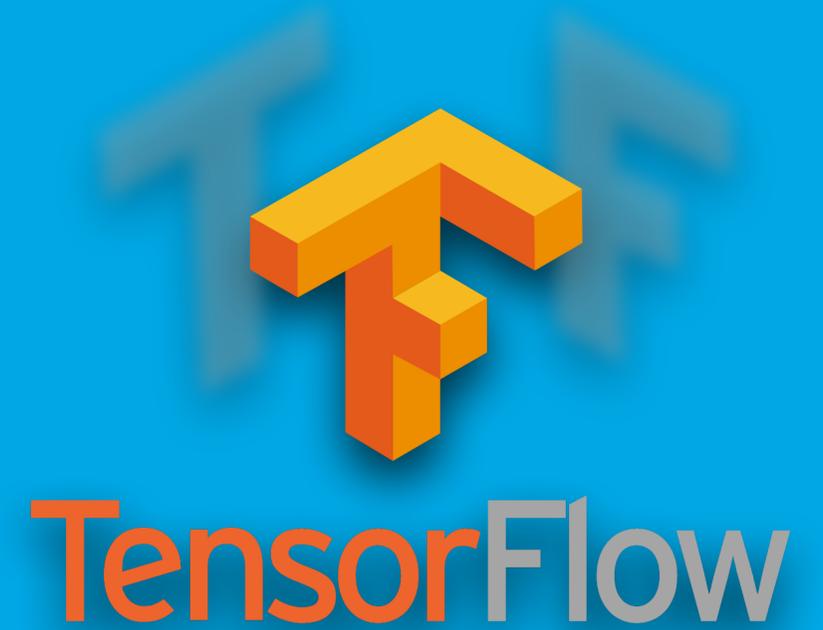
Active investors see their strategies automated while passive alternatives get cheaper



London-based start-up Havelock is currently trying to fashion an algorithm that mimics Warren Buffett. Such attempts are helping to undermine what was once a prominent feature of the investing landscape: the cult of the star stock picker. © Bloomberg



**NVIDIA®**



**Demos**

## **The Random Walk Hypothesis Revisited**

Jupyter Notebook 01

## **Discovering Statistical Inefficiencies with Machine Learning**

Jupyter Notebook 02

Jupyter Notebook 03

## **The AI Machine — AI-Powered Algo Trading**

# The AI Machine

# Algorithmic Trading

**Data**

**Features and Labels**

**Machine & Deep Learning**

**Backtesting**

**Deployment**

**EUR/USD  
exchange rate**

**time series features**

**support vector  
machine &  
DNN classifier**

**vectorized  
backtesting**

**online algorithm**

**1 minute bars**

**economic indicators**

**training & testing**

**event-based  
backtesting**

**cloud deployment**

**August & September  
2019**

**directional movement**

**validation**

**visualization**

**monitoring & risk  
management**

# Algorithmic Trading

Data

Features and Labels

Machine & Deep Learning

Backtesting

Deployment



REFINITIV

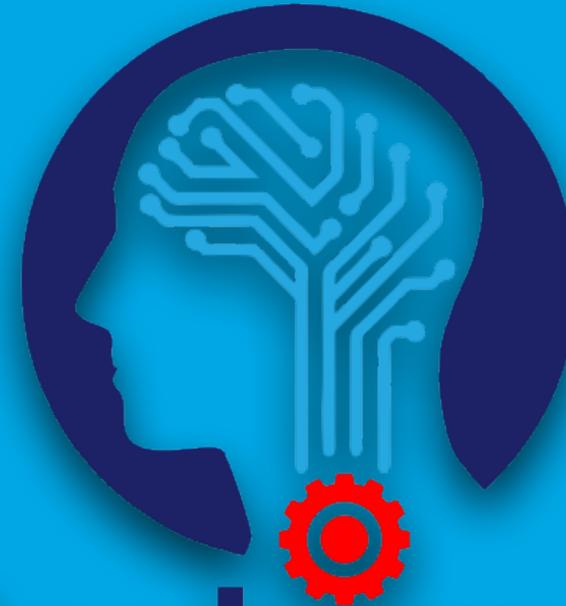


docker



# The AI Machine

*Standardized deployment  
of AI-powered  
algorithmic trading strategies.*



as of: September 17, 2019 at 8:39:06 PM GMT+7

interrupt	model attributes	total	trading account	hedge account	details								
...	s... 1 ↑	inst	fr...	model n...	position	P&L	unreal	real	position	P&L	position	P&L	...
	Toggled	EUR_USD	M30	bagging	0	4.37	0.00	4.37	0	4.37	0	0.00	

### audit details ●

model name	frequency	ML technique	lags	trade quantity	SL distance %	take profit %	accuracy	toggle wait time freq	max sl hits	current stop loss count
bagging	M30	BaggingClassifier	5	2000	0.125	0.175	83.33%	1	2	0

Zoom: 30S 1M 5M 10M 30M 1H 3H 12H all



10:12 10:30 10:42 11:00 11:12 11:30 11:42 12:00 12:12 12:30 12:42 13:00 13:12

## The AI Machine

### Prediction Engines

Technical Rules  
Statistical Methods  
Machine Learning  
Deep Learning  
Automated ML  
...

### Deployment & Execution

Data Streaming  
Transaction Streaming  
Position Management  
Profit & Loss Calculation  
Visualization  
Logging & Reporting  
...

### Decision Rules

Entry Rules  
Stop Loss  
Trailing Stop Loss  
Dynamic Stop Loss  
Take Profit  
Position Sizing  
Capital Allocation  
...

# Conclusions

1. Finance has long been driven by the “**beauty myth**” — elegant but too simplistic models, equations and approaches.
2. The availability of **big financial data** (historical—streaming, structured—unstructured) gave rise to data-driven finance.
3. It might be assumed that the “**unreasonable effectiveness of big data**” holds true in the financial domain as well.
4. Due to the availability of big data (e.g. billions of hours of virtual car driving, billions of self-played games), **Artificial Intelligence** (AI) is changing almost every area of our lives.
5. It is to be assumed that in the same way the **combination of data-driven and AI-first finance** will influence and change finance and algorithmic trading for good.

# The Python Quants The AI Machine

Dr. Yves J. Hilpisch  
+49 3212 112 9194  
tpq.io | ai@tpq.io | @dyjh

