ROBECO
The Investment Engineers

Momentum Think Tank 2023

Surfing the ocean of alternative data to reach new investment heights

Weili Zhou
Head of Quantitative Investments, Robeco
September 20, 2023

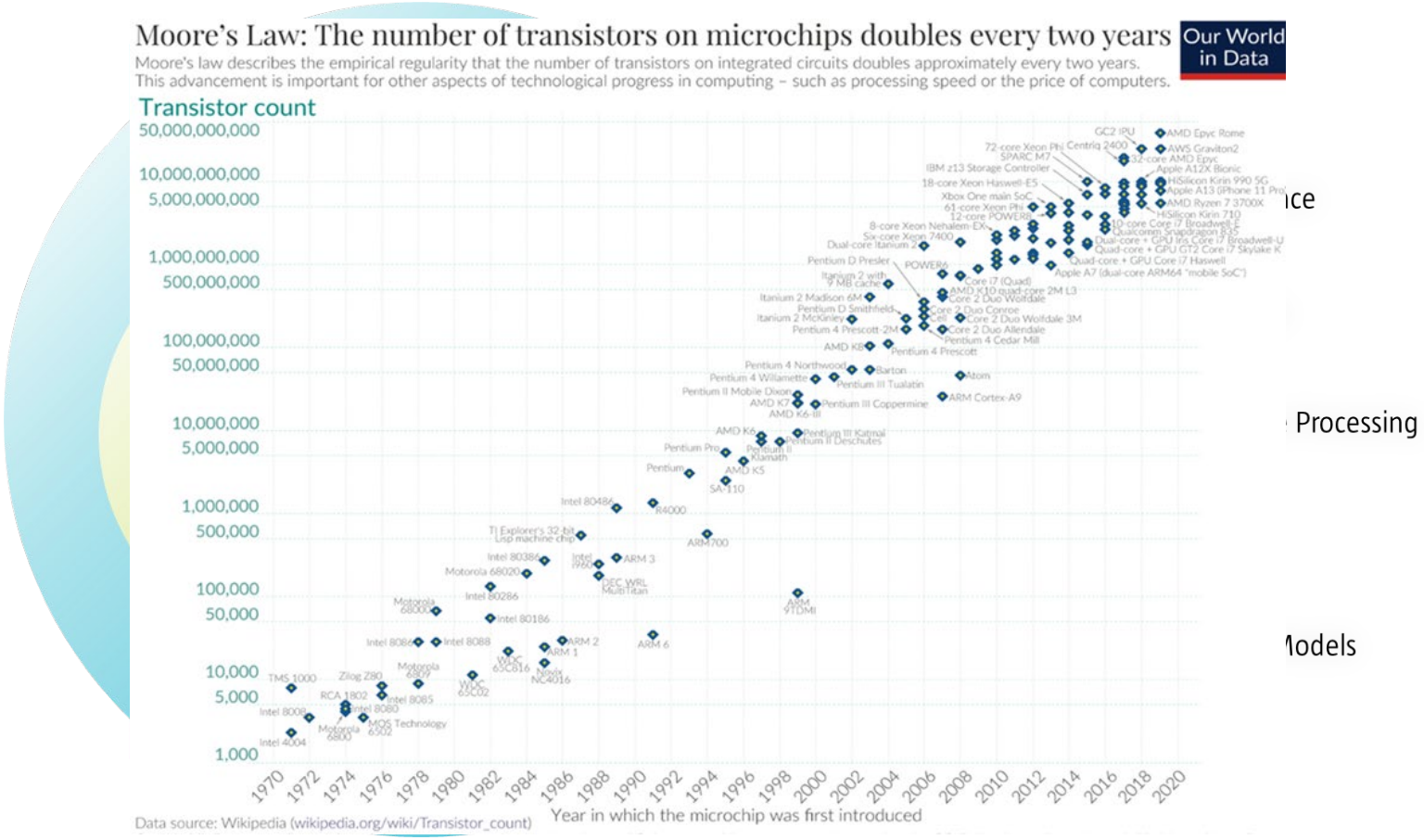
Sustainable Investing Expertise by
ROBECOSAM

Today's Agenda

- Navigating the Ocean of Alternative Data
- Machine Learning for Quant Investing
- From NLP to alpha
- Conclusion and Q&A

Setting the Context: Three Major Technology Trends Transforming the World

Data, Algorithms, and Computing Power



ice

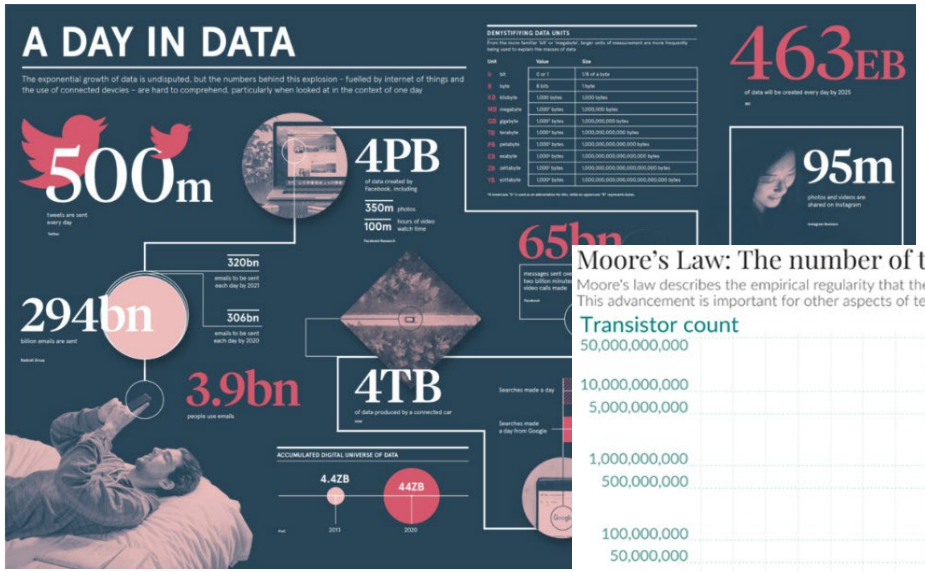
Processing

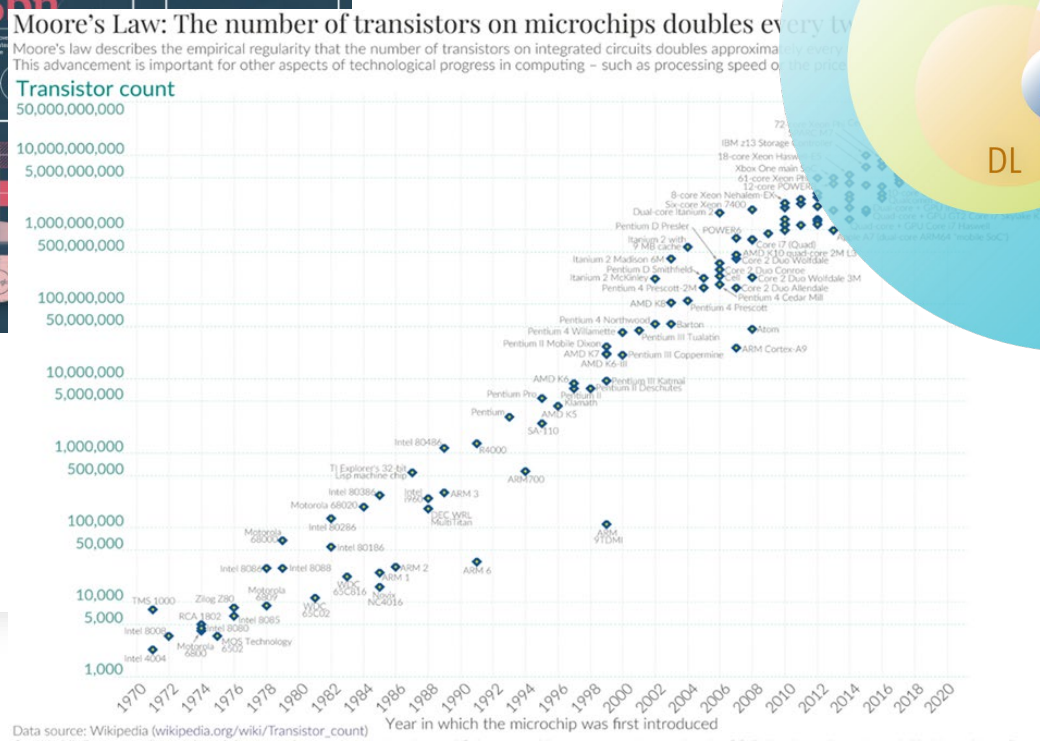
Models

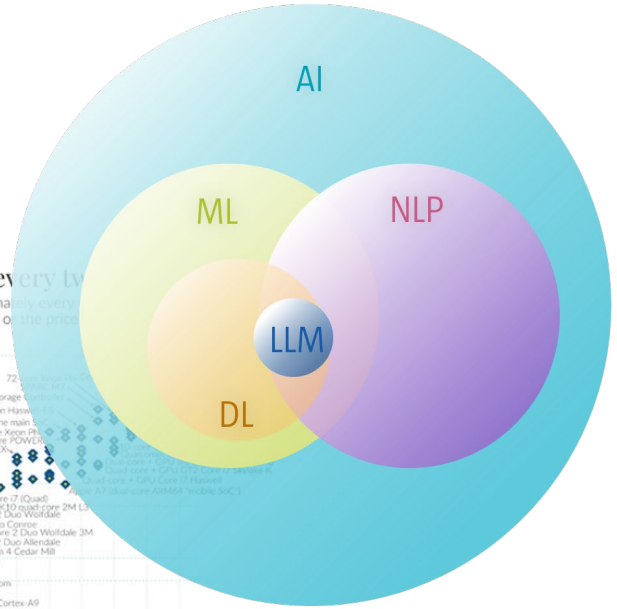
ROBECO

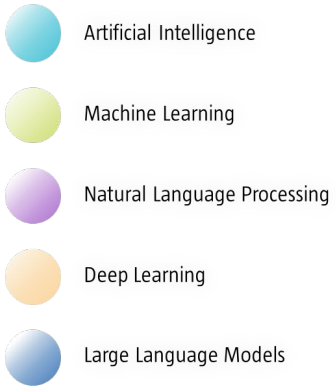
Setting the Context: Three Major Technology Trends Transforming the World

Data, Algorithms, and Computing Power

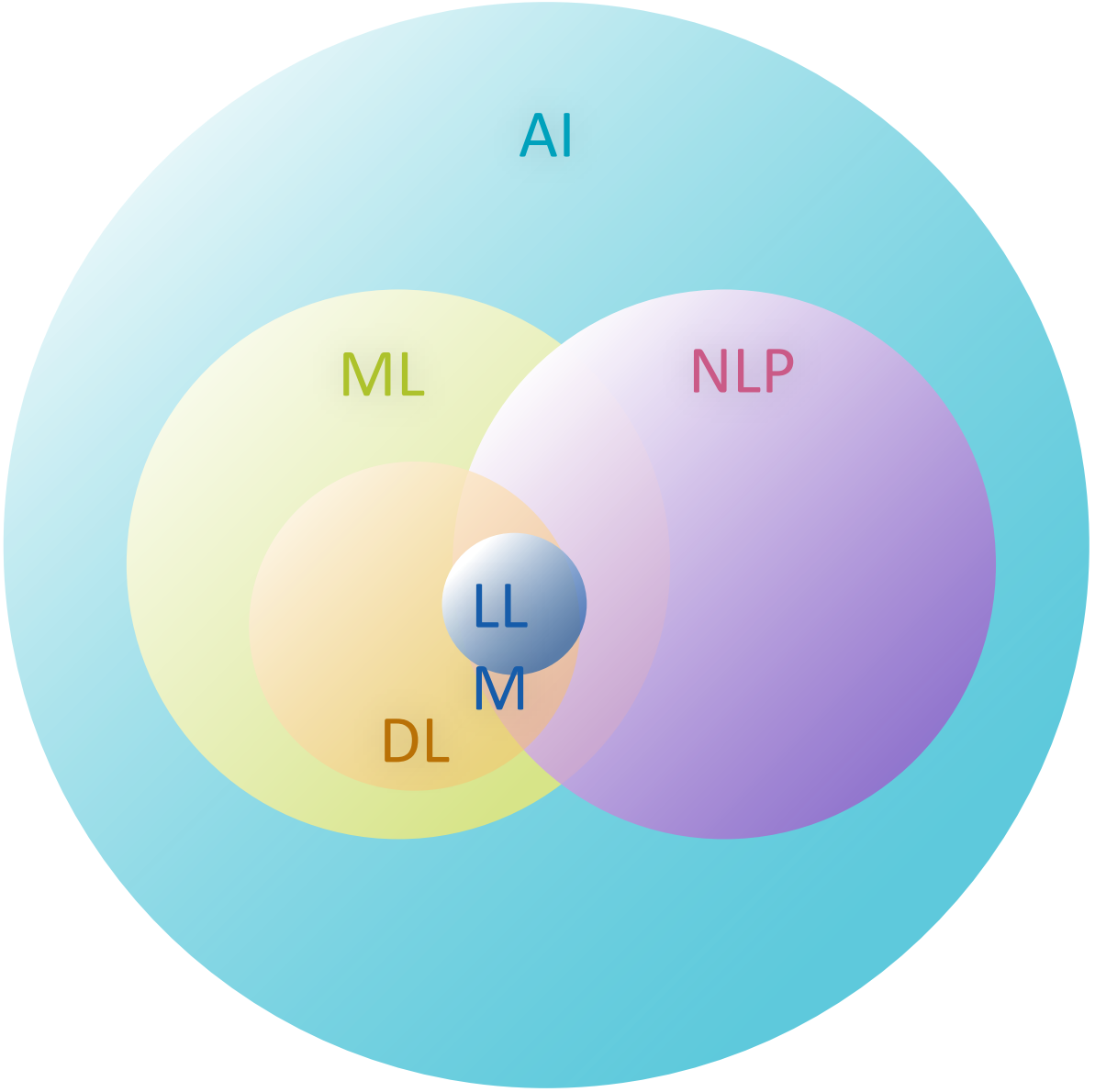








Setting the Context: The Current AI Landscape



Artificial Intelligence

Machine Learning

Natural Language Processing

Deep Learning

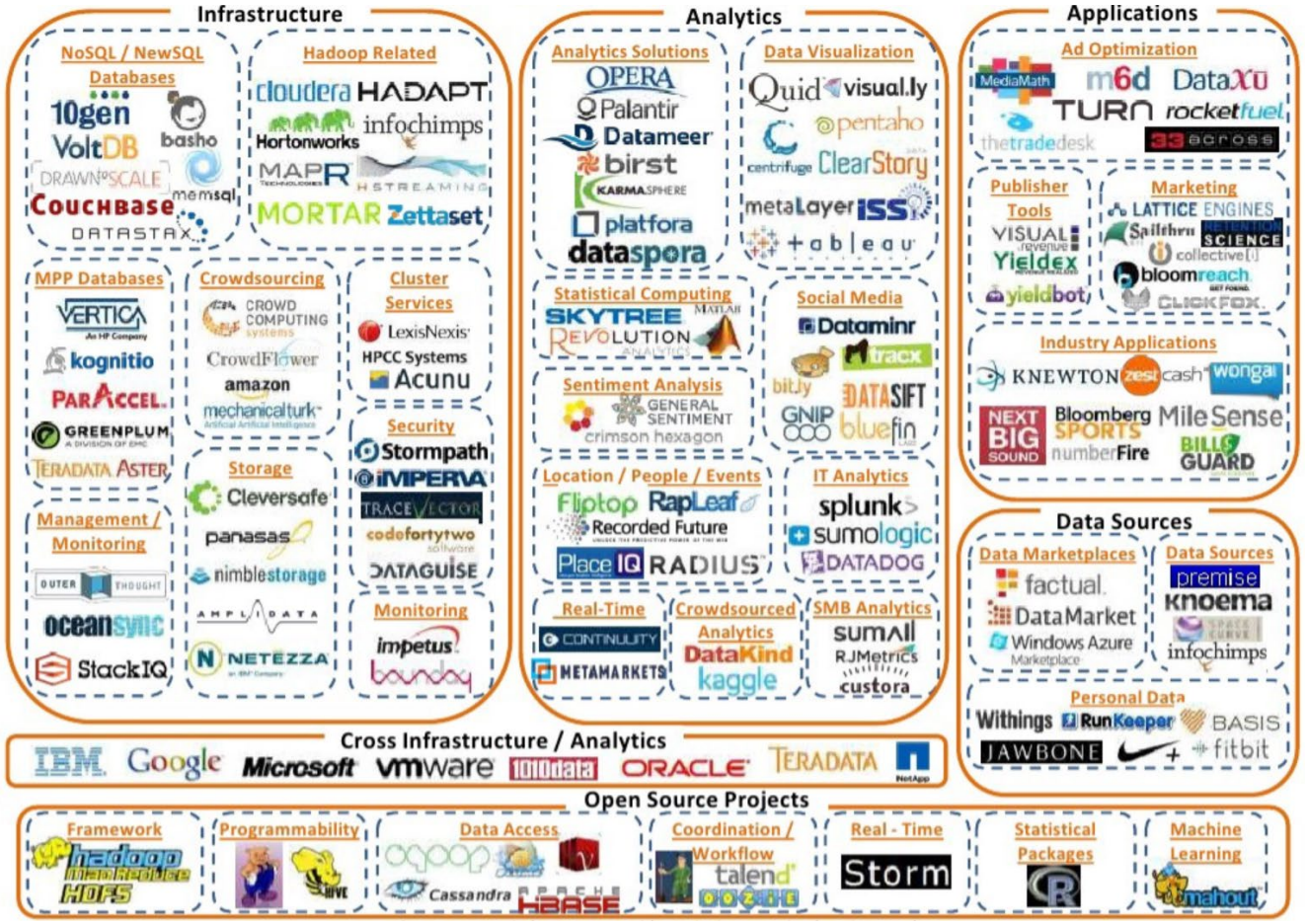
Large Language Models

Add your disclaimer here

Navigating the Ocean of Alternative Data

Alternative Data

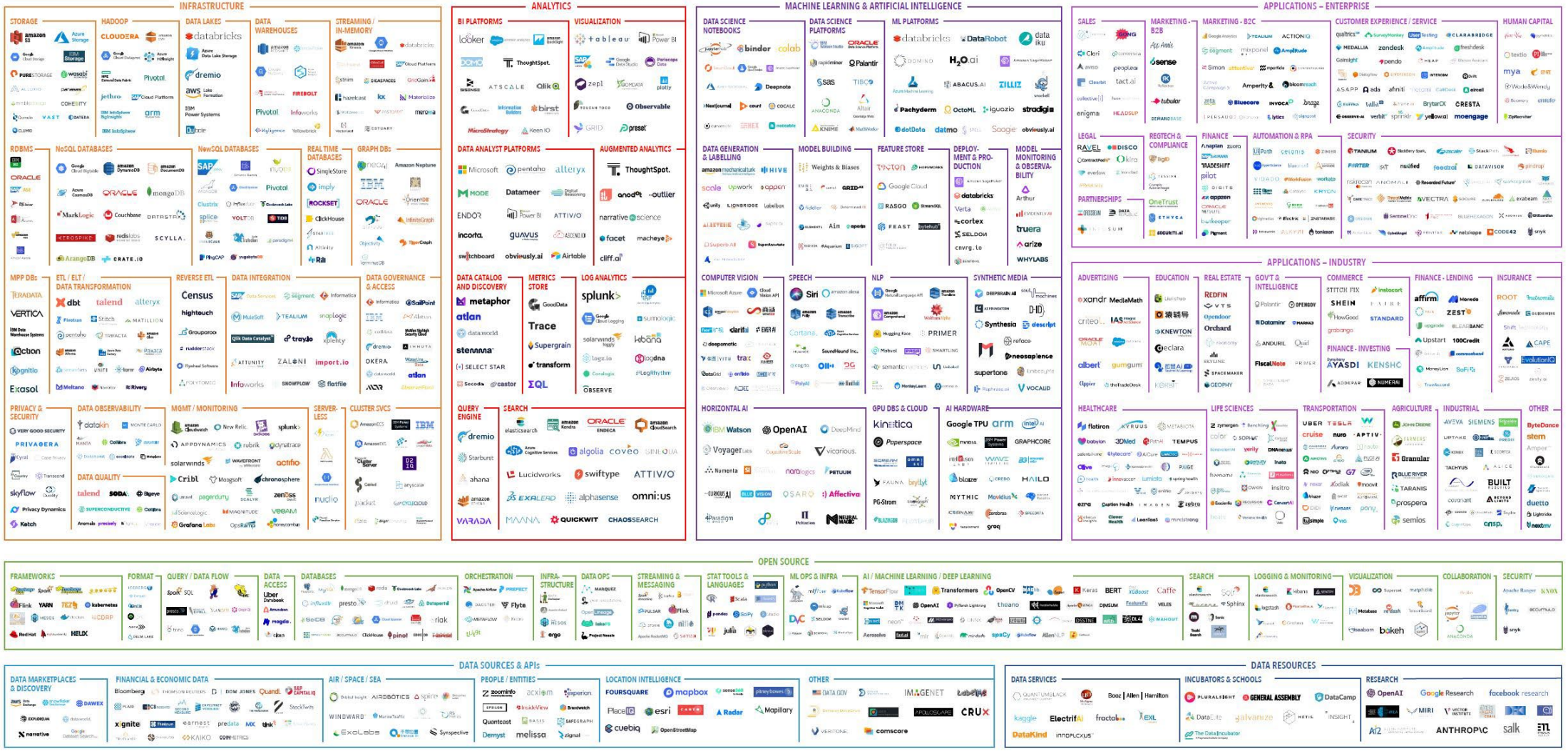
The data landscape has changed over time



Source: Matt Turck

Alternative Data

The data landscape has changed over time



Source: Matt Turck

Alternative Data is distinct from traditional data

At the same time, some alternative datasets are becoming more mainstream

Traditional Data

Market + Fundamentals

THOMSON REUTERS

STANDARD &POOR'S

Bloomberg

Alternative Data

- Not readily available
- Non-standard
- Usually very large in scale
- High potential benefits

E.g. Web-scraped + Holdings based

DOW JONES NEWSWIRES

IHS Markit

E.g. Satellite imagery + Geo-location

GAMAYA

ADVAN
Research through Technology

Complexity + Opportunity

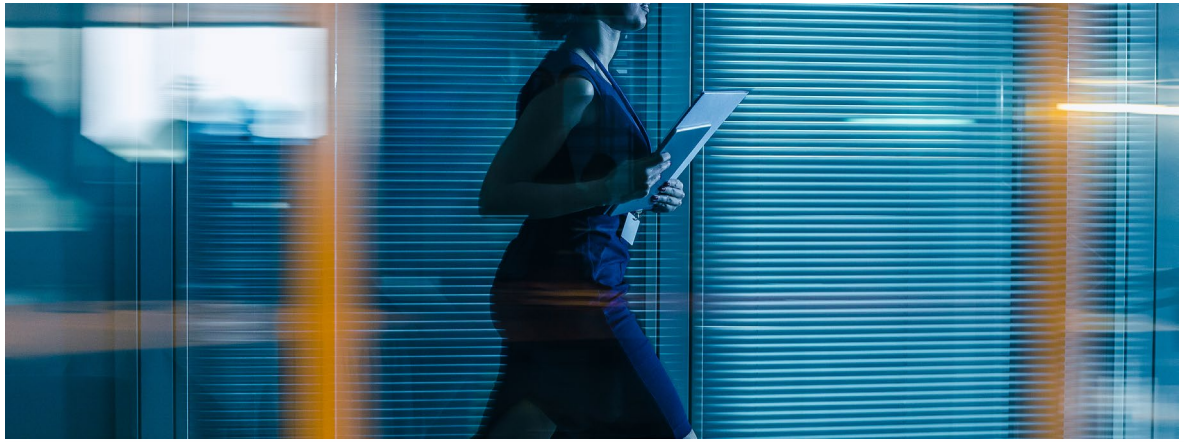
Crowdedness

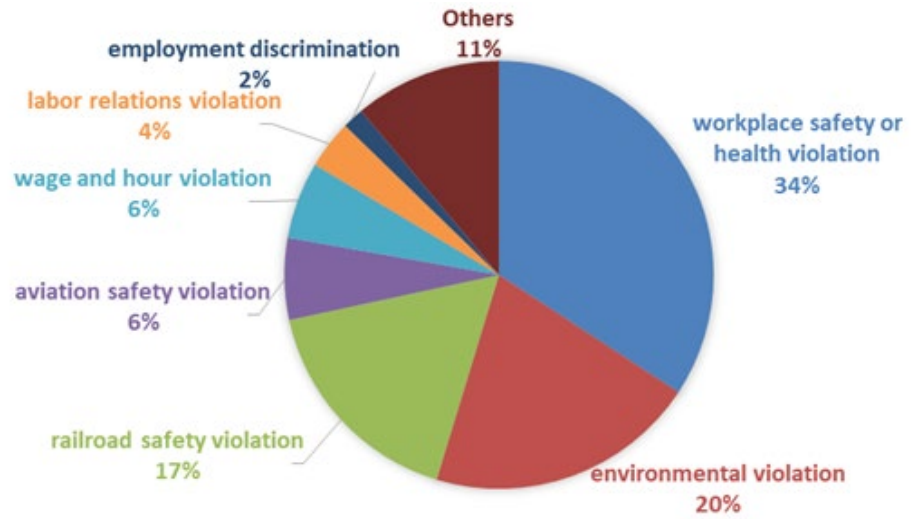
Example I: Corporate Violation and Penalty

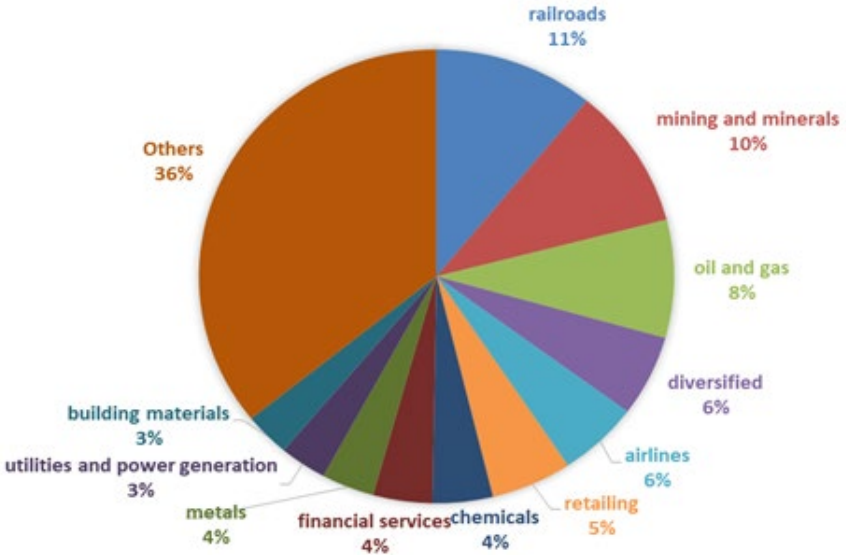
Based on Government data collected by an NGO

> Sensibility: the announcement of a regulatory violation is an exceptionally clean signal to the market about the extent to which the firm in question abides by its legal obligations (Armour et al., 2017)

> Excessive violations and penalties versus industry peers can possibly indicate not only lax governance structure and low management quality, but possibly also ineffective/risk-prone corporate culture

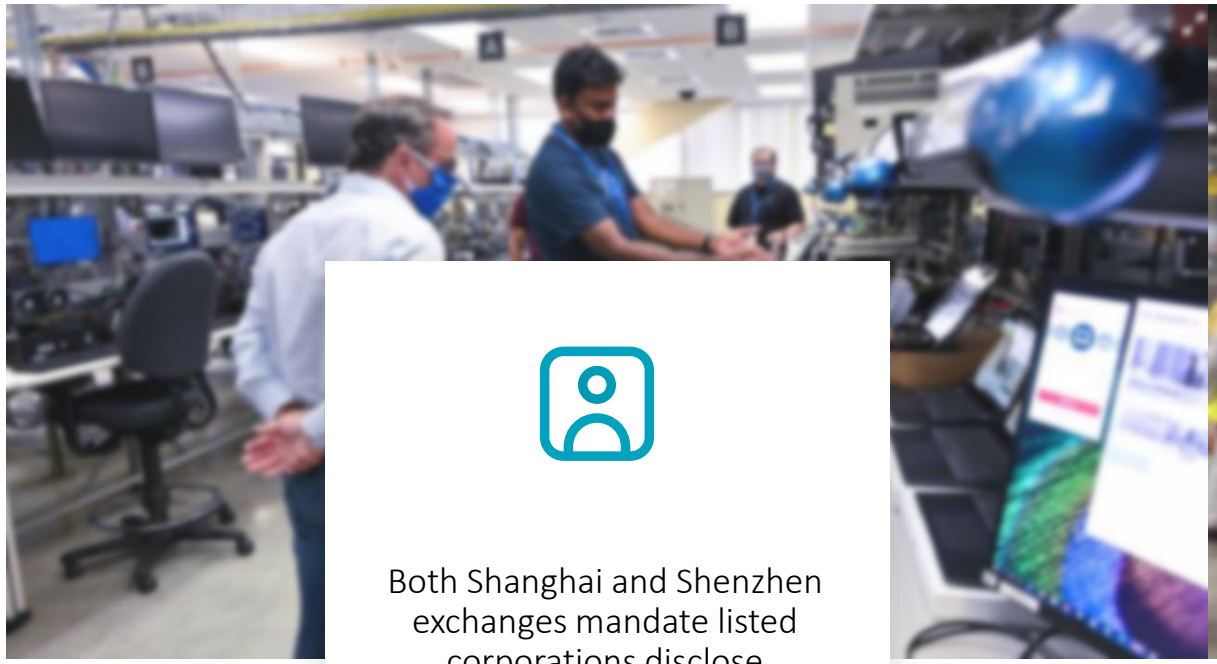




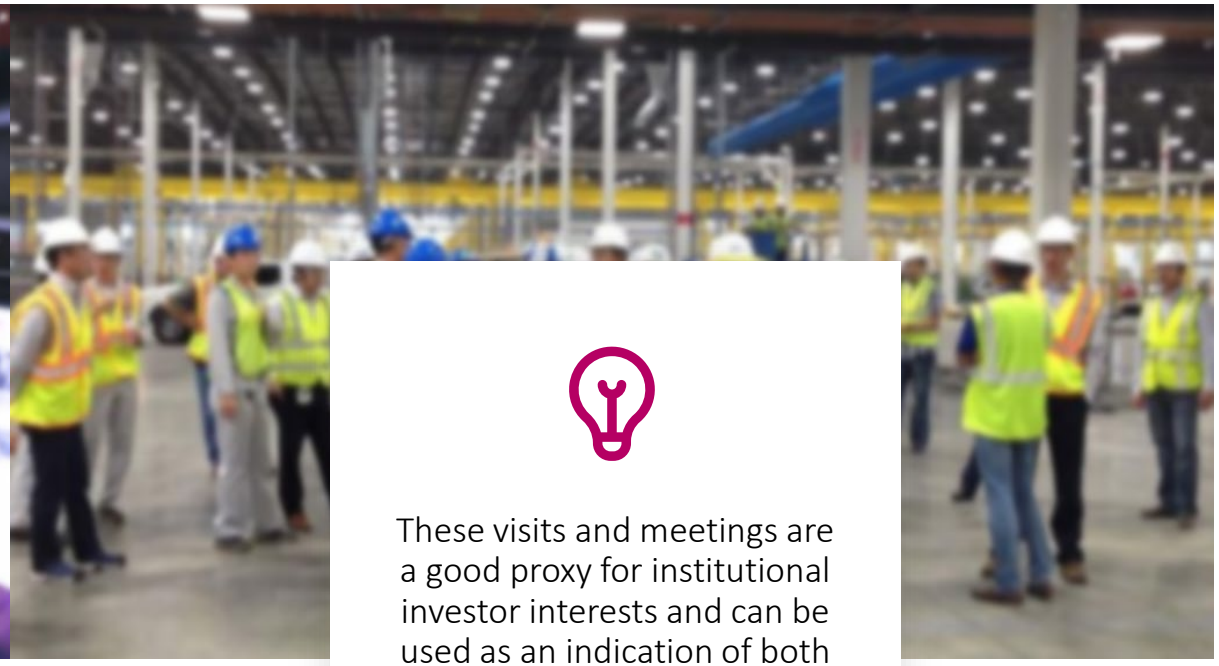


Example II: Institutional Investor Site Visits

Unique data particular to the Chinese A-shares market



Both Shanghai and Shenzhen exchanges mandate listed corporations disclose company site visits



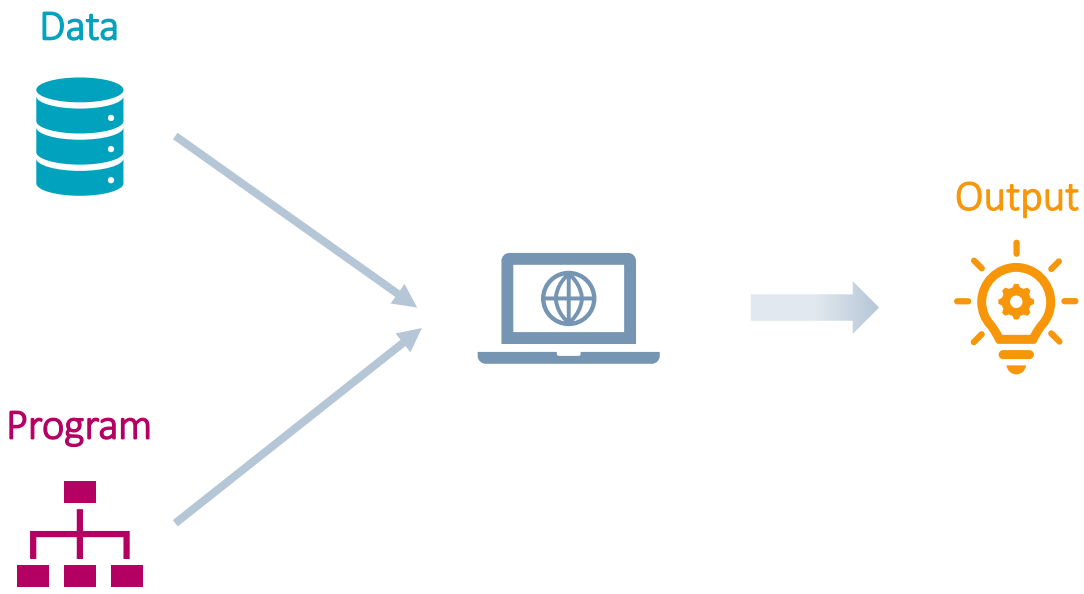
These visits and meetings are a good proxy for institutional investor interests and can be used as an indication of both corporate outperformance (stock-selection) and industry aggregate views (industry-timing)

Machine Learning for Quant Investing

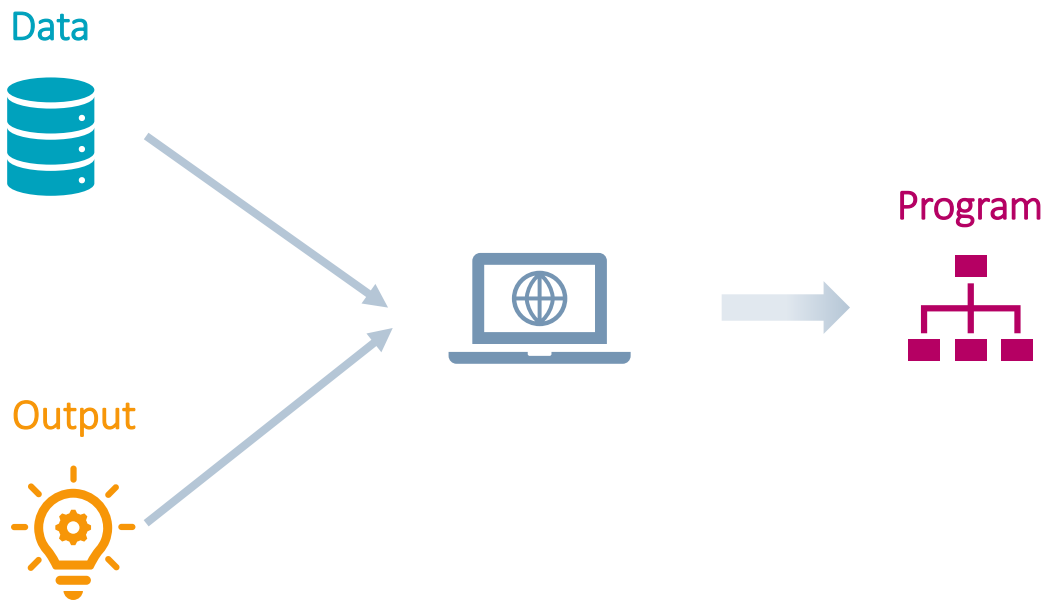
What is machine learning?

"Machine learning algorithms improve automatically through experience and by the use of data."

Traditional programming



Machine Learning



Advantages of Machine Learning techniques

Data driven

- Identify most relevant features
- Efficient research cycle

Flexible

- Allow for nonlinear relations
- Allow for interaction effects

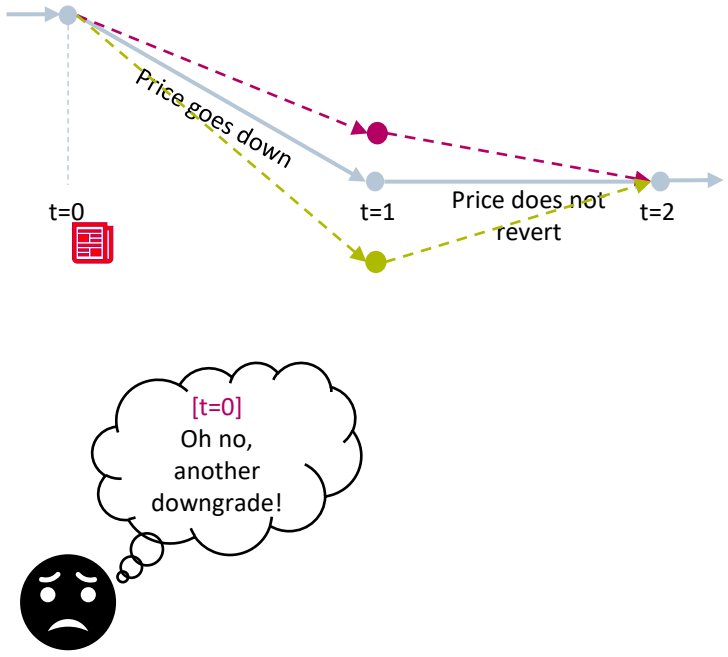
Forward-looking

- Prevent lookahead bias
- Deduce structural patterns

Example III: Predicting returns using Machine Learning

News and return reversal interaction

News changes the inherent value of a security



Predicted Return over the next 10 days

Return over past 10 days	Abnormal News Count over the past 10 days: Less Abnormal News			No Change			More Abnormal News
Negative Return	16%	13%	10%	8%	5%	2%	-1%
	11%	9%	7%	5%	3%	1%	-1%
	6%	5%	4%	3%	1%	0%	-1%
No Change	1%	0%	0%	0%	0%	0%	-1%
	-4%	-4%	-3%	-3%	-2%	-1%	-1%
	-10%	-8%	-7%	-5%	-3%	-2%	0%
Positive Return	-15%	-12%	-10%	-8%	-5%	-3%	0%

From NLP to Alpha

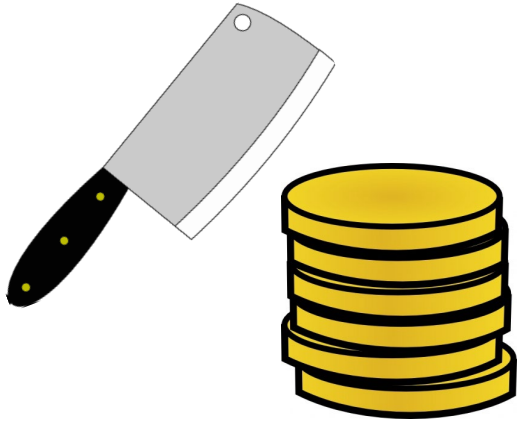
NLP

Translating text to mathematical objects

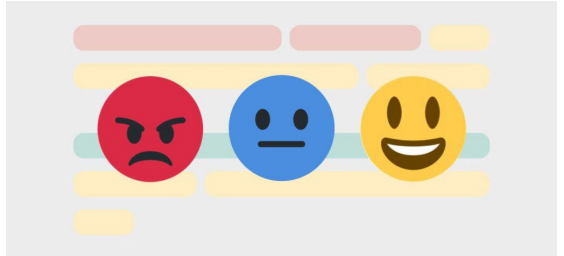
> The general NLP process

Gathering and cleaning textual data

Tokenize textual data



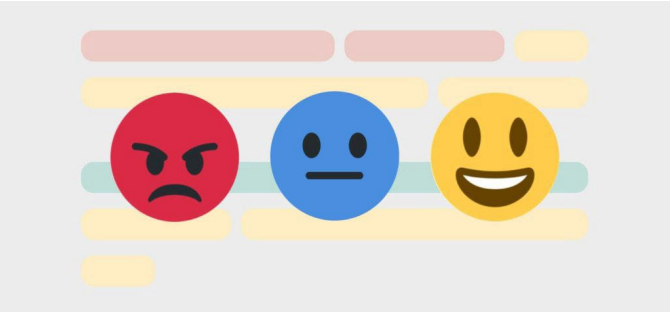
Translate tokens to application



NLP applications

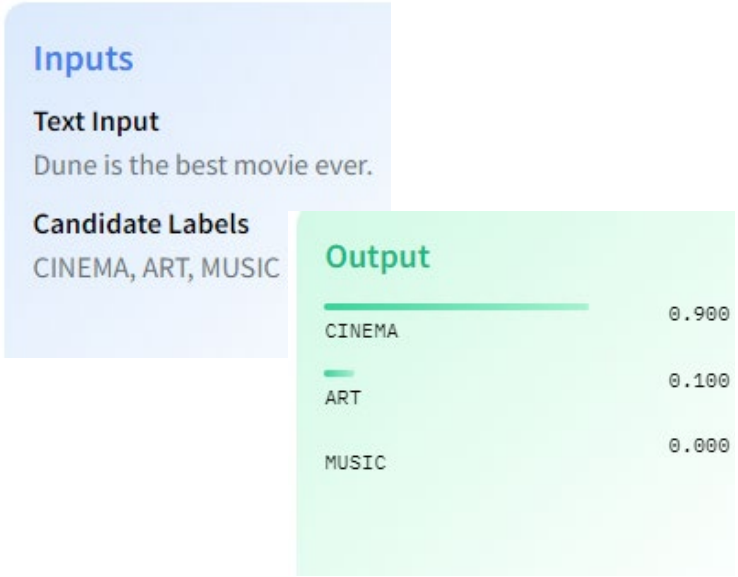
Sentiment classification

- Identify text tone



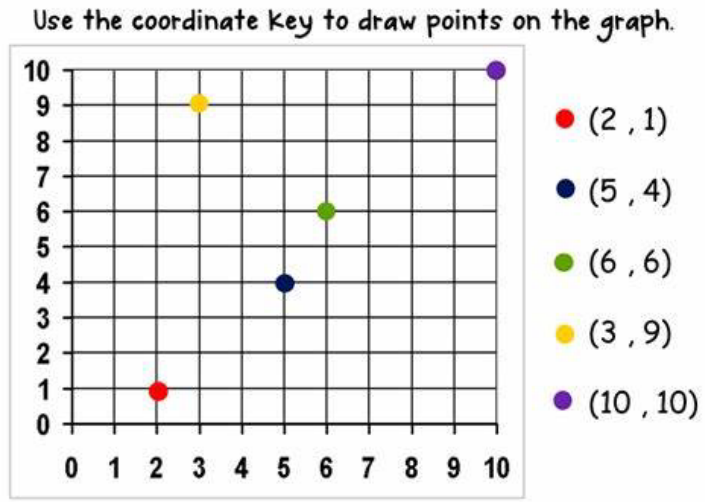
Zero-Shot classification

- Train a model on some classes
- Predict a class the model does not know



Sentence similarity

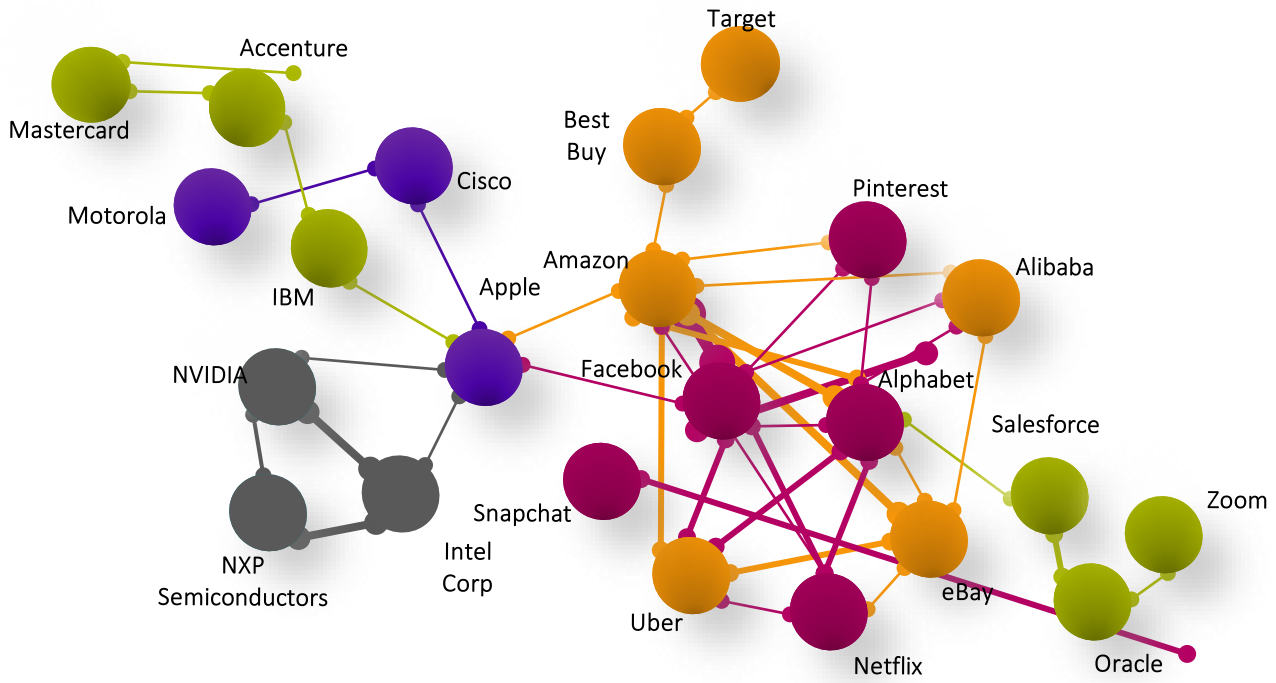
- Translate sentences to coordinates
- Calculate distance between coordinates



Example IV: Industry classification

- GICS classification is intuitive but limited
- ML techniques can incorporate a lot more information and make more realistic peer groups
- Optimal peer identification combines human understanding with data-driven grouping

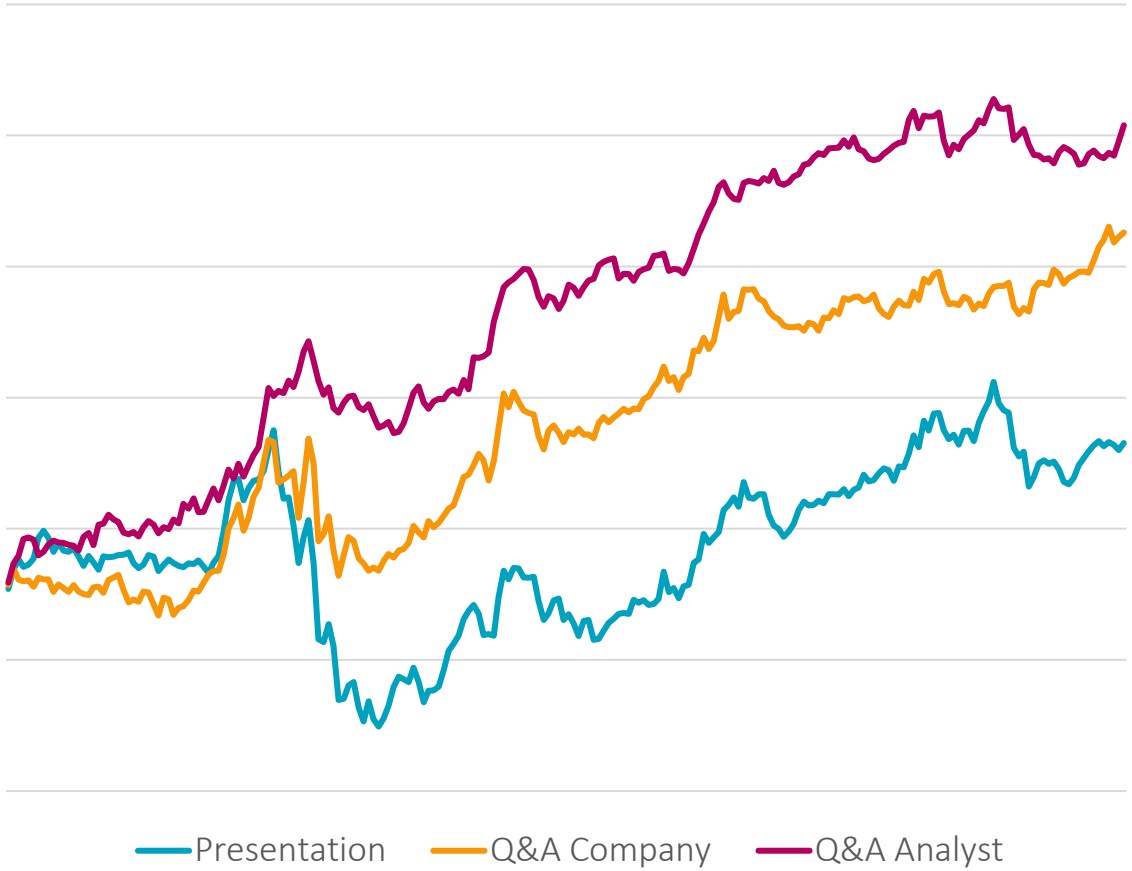
Interactive Media	Retail
Google	amazon
Broadcasting	Hardware
NETFLIX	



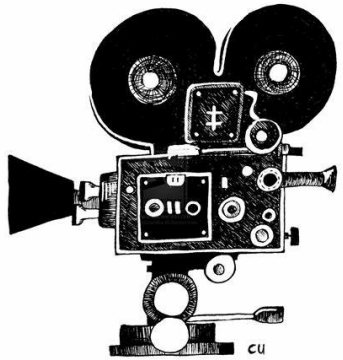
Source: Robeco, Refinitiv. For illustration only.

Example V: The value of an earnings call

Cumulative return



Takeaways



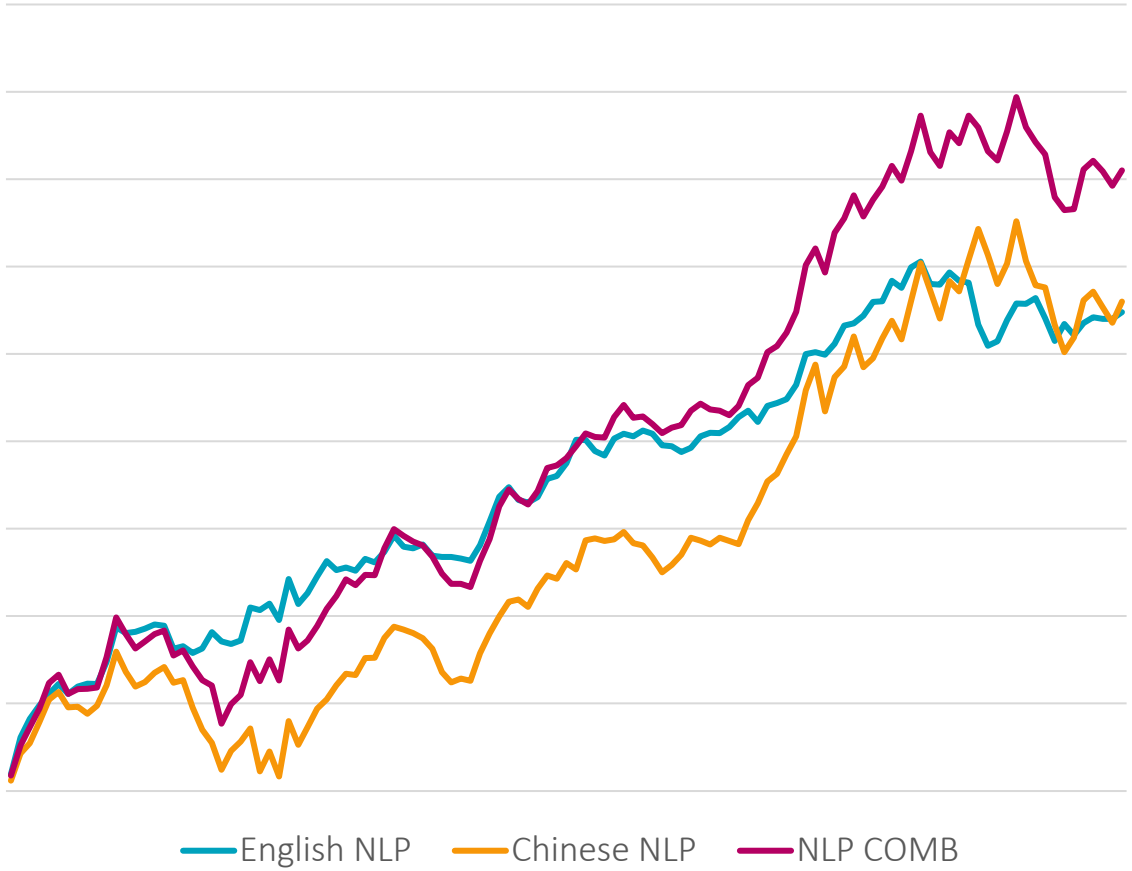
<

Management < Analysts

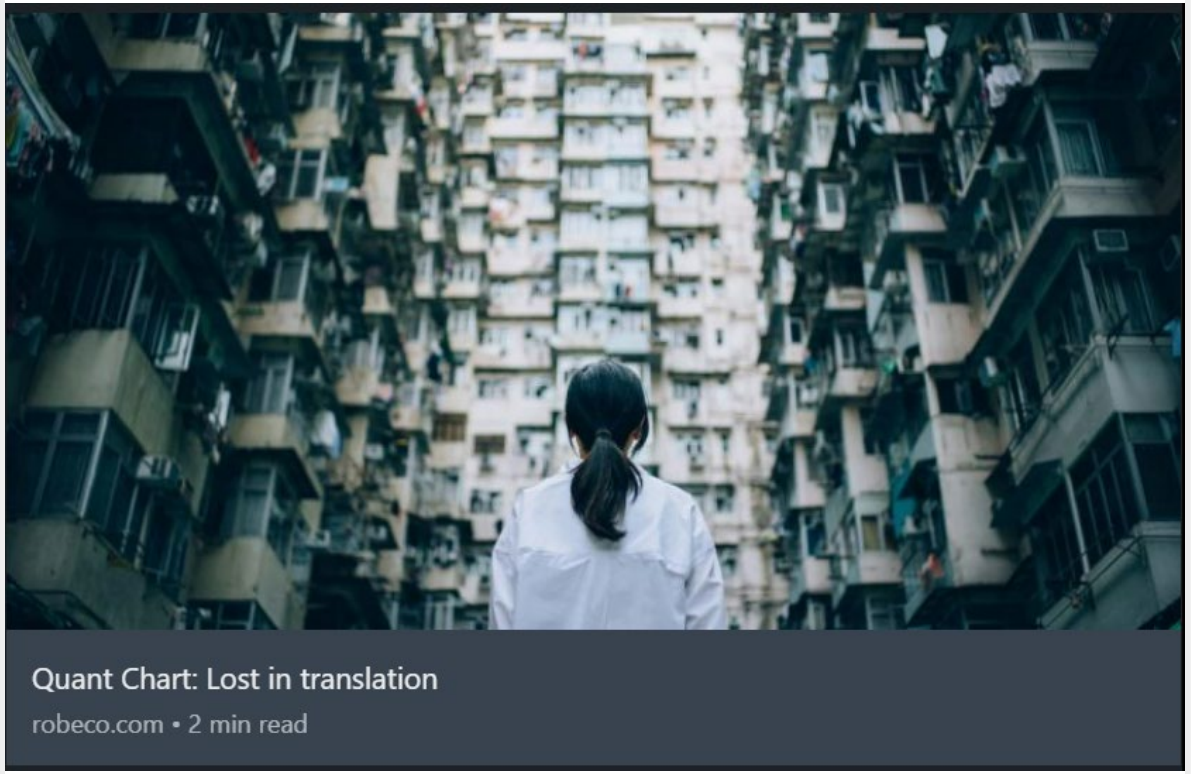
Example VI: Lost in translation

The underlying language matters

Cumulative return



LinkedIn post



Conclusion and Q&A

Conclusion

- Like in many other industries, big data and AI (next-gen quant applications) is here to stay in asset management
- Although most natural for quant investing, next-gen quant applications can also be of great help in fundamental investing
  - Quant: next-gen quant can be used to detect non-linear patterns, estimate sentiments and trending topics, etc.
  - Fundamental: next-gen quant applications can be used to summarize research reports, suggest investment ideas, (mostly) automate report writing, etc.

- Next-gen quant is not a panacea, but it can tremendously augment existing capabilities and increase efficiency
- At the same time, it is still very important to keep “human-in-the-loop”, especially for AI applications
- The technology may seem daunting, but the barrier to entry is being lowered continuously. At the same time, many firms are racing ahead in their own exploration. ***The time to start is today!***

For Further Reading

1. ON MACHINE LEARNING APPLICATIONS IN INVESTMENTS

Mike Chen, PhD
Head, Alternative Alpha Research, Robeco

Weili Zhou, CFA
Head, Quant Equity Research, Robeco

Introduction

In recent years, machine learning (ML) has been a popular technique in various domains, ranging from streaming video and online shopping recommendations to image detection and generation to autonomous driving. The attraction and desire to apply machine learning in finance are no different.

- "The global AI fintech market is predicted to grow at a CAGR of 25.3% between 2022 and 2027" (Columbus 2020).
- "A survey of IT executives in banking finds that 85% have a 'clear strategy' for adopting AI in developing new products and services" (Nadeem 2018).

Putting aside the common and widespread confusion between artificial intelligence (AI) and ML (see, e.g., Cao 2018; Nadeem 2018), the growth of ML in finance is projected to be much faster than that of the overall industry itself, as the previous quotes suggest. Faced with this outlook, practitioners may want answers to the following questions:

- What does ML bring to the table compared with traditional techniques?
- How do I make ML for finance work? Are there special considerations? What are some common pitfalls?
- What are some examples of ML applied to finance?

In this chapter, we explore how ML can be applied from a practitioner's perspective and attempt to answer many common questions, including the ones above.[1]

The first section of the chapter discusses practitioners' motivations for using ML, common challenges in implementing ML for finance, and solutions. The second section discusses several concrete examples of ML applications in finance and, in particular, equity investments.

Motivations, Challenges, and Solutions in Applying ML in Investments

In this section, we discuss reasons for applying ML, the unique challenges involved, and how to avoid common pitfalls in the process.

Motivations

The primary attraction of applying ML to equity investing, as with almost all investment-related endeavors, is the promise of higher risk-adjusted return. The hypothesis is that these techniques, explicitly designed for prediction tasks based on high-dimensional data and without any functional form specification, should excel at predicting future equity returns.

Emerging academic literature and collective practitioner experience support this hypothesis. In recent years, practitioners have successfully applied ML algorithms to predict equity returns, and ML-based return prediction algorithms have been making their way into quantitative investment models. These algorithms have been used worldwide in both developed and emerging markets, for large-cap and small-cap investment universes, and with single-country or multi-country strategies.[2] In general, practitioners have found that ML-derived alpha models outperform those generated from more traditional linear models[3] in predicting cross-sectional equity returns.

[1]Readers interested in the theoretical underpinnings of ML algorithms, such as random forest or neural networks, should read Hastie, Tibshirani, and Friedman (2009) and Goodfellow, Bengio, and Courville (2016).

[2]There are also numerous academic studies on using ML to predict returns. For example, ML techniques have been applied in a single-country setting by Gu, Kelly, and Xiu (2020) to the United States, by Abe and Nakayama (2018) to Japan, and by Leippold, Wang, and Zhou (2022) to China's A-share markets. Similarly, in a multi-country/regional setting, ML has been applied by Tobek and Hronec (2021) and Leung, Lohre, Mischlich, Shea, and Stroh (2021) to developed markets and by Hanauer and Kalsbach (2022) to emerging markets.

[3]For linear equity models, see, for example, Grinold and Kahn (1999).

- Generative models like GPT could revolutionize the industry
- These models offer potential benefits while also presenting challenges
- Asset management can adapt and embrace GPT while also addressing limitations

Since its launch on 30 November 2022, ChatGPT[1] has become the fastest application to reach 100 million monthly active users[2,3] The excitement that the model has generated in the public's imagination has led to wild predictions about its impact on all aspects of society, ranging from the overly optimistic to the profoundly pessimistic. Less than six months after ChatGPT's launch, GPT-4 was released. As the time of writing, GPT-4 appears to be significantly more powerful than ChatGPT. So, what does it all mean?

In this article, we take a clear-eyed look at some possible implications of GPT, and more broadly, generative models, on the asset management industry. We give examples of how GPT might be applied in various workflows common to the asset management industry, and how asset owners and managers could position themselves to benefit as much as possible from the anticipated future generative model evolution. We begin our discussion by briefly looking back on the development of natural language processing (NLP) models that have led to Generative Pre-trained Transformers (GPT).

A quick survey on NLP models

NLP is a subfield of machine learning (ML) that focuses on enabling computers to understand, interpret, and generate human language. It combines computational techniques with linguistic knowledge to do this in a way that mimics human understanding and communication. Since the first NLP model in the 1960s[4], leaps in computational power particularly in the last three decades have meant that the ability to analyze, process and interact with text on an ever-larger scale has increased tremendously.

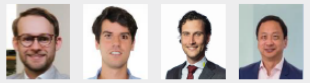
NLP applications for financial investments in particular started in 2007 with the humble bag-of-words (BoW) approach[5], which is based on identifying words associated with either positive or negative sentiment as listed in a dictionary. Despite the simple nature of the model, a BoW approach to basic sentiment detection (overall positive or overall negative) works surprisingly well.

[1] Apart from the examples given, none of this paper was written by ChatGPT or any other form of AI.
[2] https://aibusiness.com/nlp/ubs-chatgpt-is-the-fastest-growing-app-of-all-time
[3] ChatGPT sets record for fastest-growing user base - analyst note | Reuters
[4] ELIZA, developed by Joseph Weizenbaum
[5] See Tetlock (2007) and Loughran and McDonald (2011)

ARTICLE 3 APRIL 2023

Marketing material for professional investors, not for onward distribution

From left to right: **Clint Howard** Quant researcher, **Tim Vogel** Quant researcher, **Thom Marchesini** Quant researcher, **Mike Chen** Head of Alternative Alpha Research
The authors would like to thank David Blitz and the Robeco Quant Equity Research Team for fruitful suggestions and discussions.



Q&A