

AI and Fintech Research

Sean S. Cao

University of Maryland


Brief Intro

- Beginning: a computer scientist (seven years), developing optimizations for compilers of programming languages.
- Transition: a Ph.D. in accounting.
- Research trajectory :

Information disclosure: production and consumption


Earlier Stage

Using fundamental data


- 
- Help investors' decision-making
 - Help corporate managers' decision-making

Current Stage

Using AI and alternative data

- 
- Help investors' decision-making
 - Help corporate managers' decision-making

Potential Future

- 
- Help household/retailers' decision-making
 - Help startup entrepreneurs' decision-making

Stage I: Fundamental research for investors

- Reexamining Growth Effects: Are All Types of Asset Growth the Same?
 - To distinguish healthy growth firms from empire-building firms through the dissection using balance sheet metrics.
- Skin in the Game: Operating Growth, Firm Performance, and Future Stock Returns.
 - Innovatively revise the imputation of balance-sheet metrics and design a trading strategy that generates significant annual and short-window returns.
- Earnings Volatility, Post–Earnings Announcement Drift, and Trading Frictions
 - To quantify useful information for investors from income statements that predicts firms' future operating performance but is not understood by investors timely.

Transition to Stage II: using AI and Alternative data

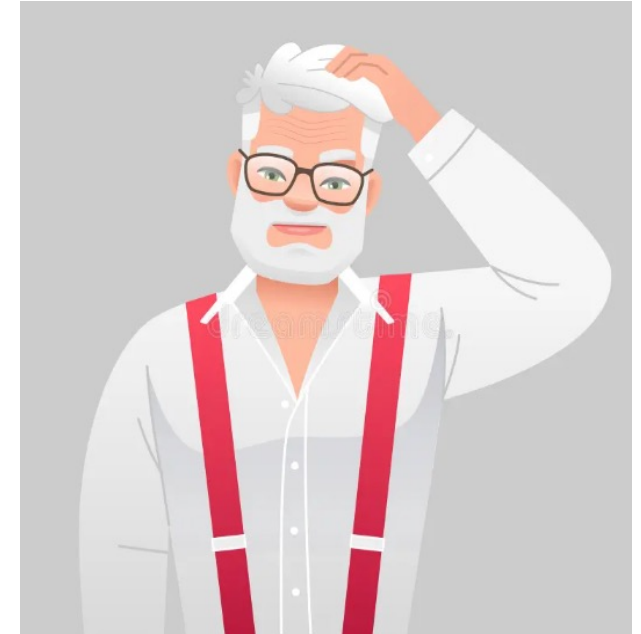
Challenges: Data does not yield a good paper; however, a good idea will.



Pondering, wondering and aging...



How to harvest good ideas by using emerging tech?

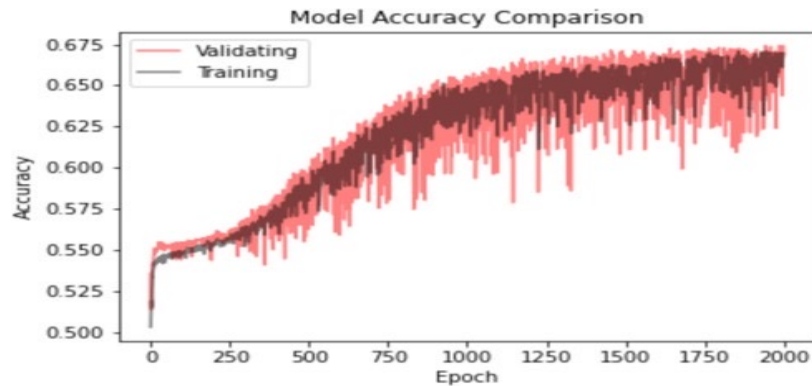


- Challenge 1: The hurdle lies not in technological skills (e.g. LLM) but good ideas.
- Challenge 2: Interpreting technology instead of leaving it as a "black box".

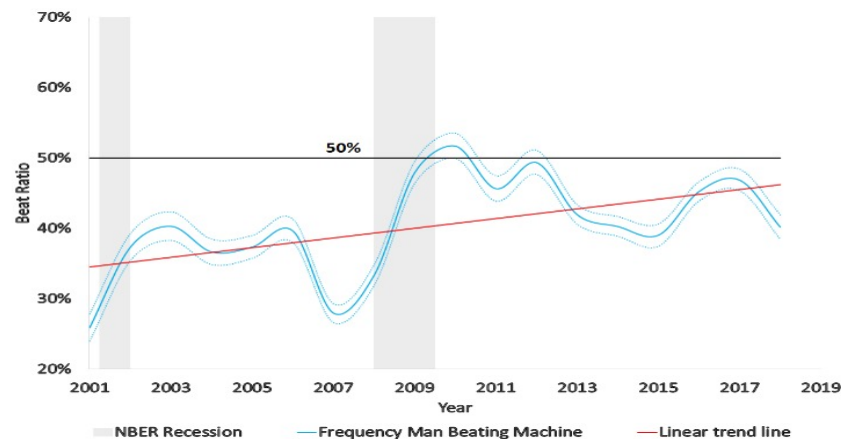
Stage II: Related AI research

□ 1. Man + Machine

- An AI stock analyst (Cao, Jiang, Wang and Yang JFE 2024)
 - **replace** human intelligence?
 - **augment** human intelligence?

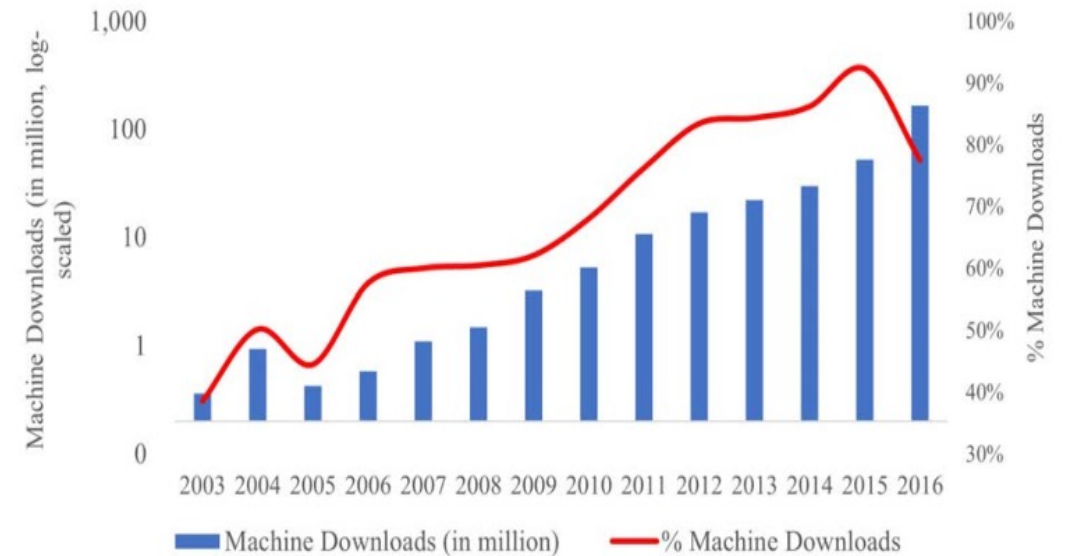


The performance of Man vs. Machine



□ 2. Machine reading news

- Machine readers are on the rise and have become significant consumers of financial reporting. (Cao, Jiang, Yang and Zhang RFS 2023)

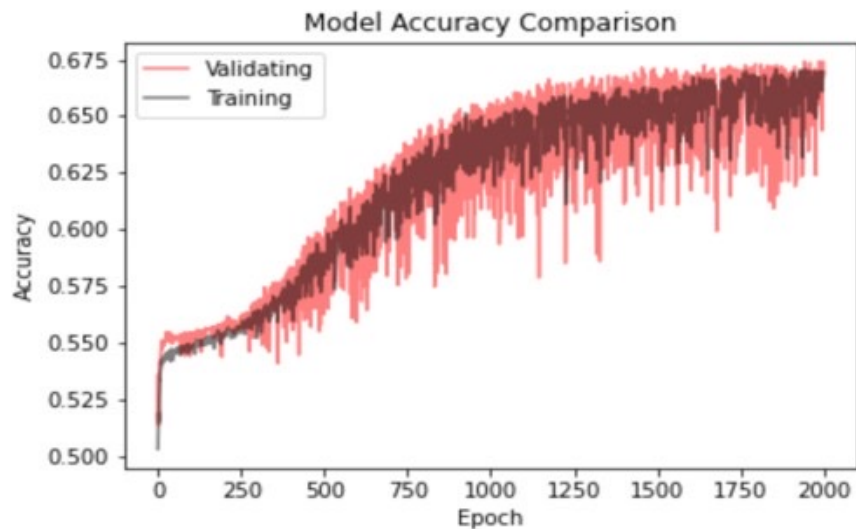


Michael J. Brennan Award for the best paper published in *Review of Financial Studies* (with coauthors), [How to Talk When Machines are Listening: Corporate Disclosure in the Age of AI](#)

Stage II: Related AI research

□ 1. Man + Machine

- An AI stock analyst (Cao, Jiang, Wang and Yang JFE 2024)
 - **replace** human intelligence?
 - **augment** human intelligence?



□ This paper has three parts:

- Machine: Can a machine beat most human analysts?
- Man: What are the human analyst advantages in the age of AI?
- Man + Machine: How can AI-equipped analysts beat machine?

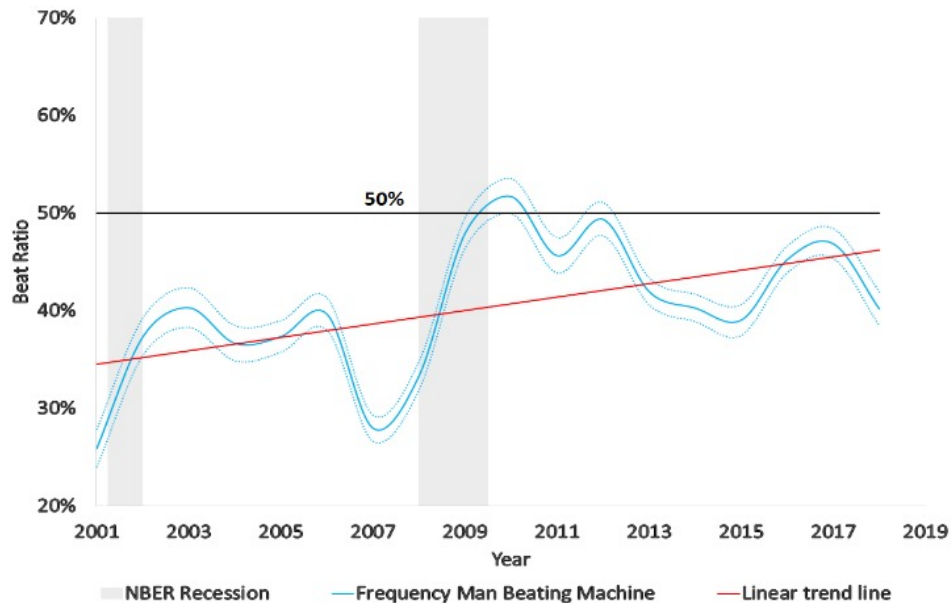
□ Variables as inputs to machine learning

- Firm Characteristics: Stock past price, past returns, past earnings, and firm characteristics known to be related to stock prices.
- Industry Variables: Industry competition, product market fluidity, industry affiliation, Industry size, etc.
- Macro variables: Industrial Production Index, Consumer Price Index, Crude Oil Price (WTI), 3-month treasury bill rate, 10-Year treasury constant-maturity rate, The BAA-AAA yield spread, etc.
- Unstructured alternative data: 10-K, 8-K, conference call transcript, Twitter, production images, etc.

Stage II: Related AI research

□ 1. Man + Machine

- An AI stock analyst (Cao, Jiang, Wang and Yang JFE 2024)
 - **replace** human intelligence?
 - **augment** human intelligence?



The performance of Man vs. Machine

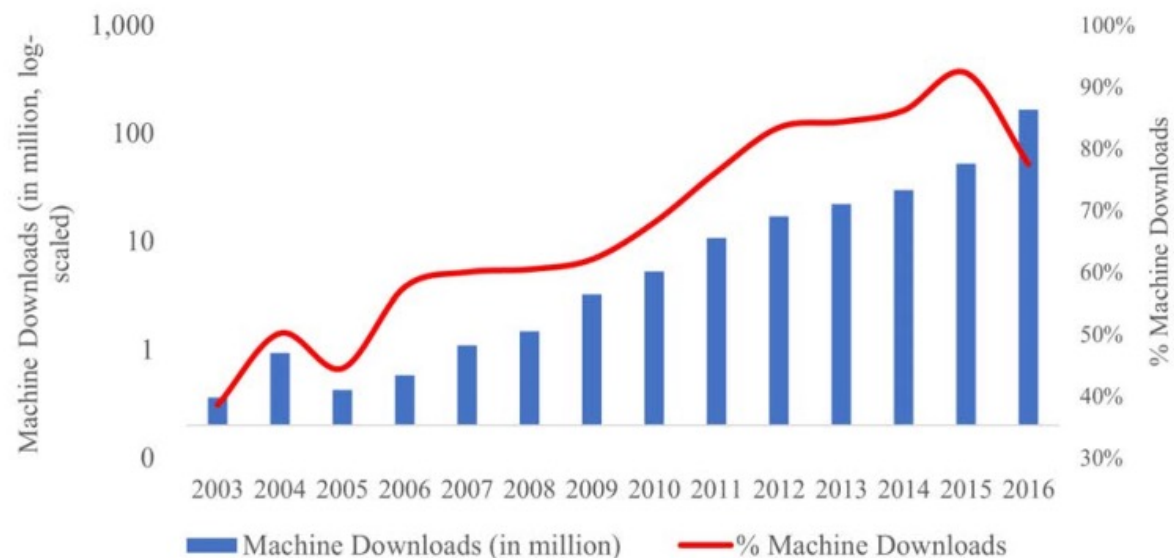
- *Profitable trading strategy*: Following AI predictions, there is an approximate **1% monthly alpha** after adjusting for risk factors.
- Who thrives in the era of AI?
 - Talents with strong reasoning or logical thinking skills to understand and analyze low-frequency events.
 - Talents who effectively leverage AI in their work; AI will augment their capabilities, rather than replacing them.
 - Talents who understand human (corporate management teams) and other soft skills like innovation.
- In conclusion, humans possess skills that machines cannot replace → The "Man + Machine" work mode surpasses both "Machine alone" and "Man alone."

Stage II: Related AI research

□ 2. Machine reading news

- Machine readers are on the rise and have become significant consumers of financial reporting.

(Cao, Jiang, Yang and Zhang RFS 2023)



Michael J. Brennan Award for the best paper published in *Review of Financial Studies* (with coauthors), [How to Talk When Machines are Listening: Corporate Disclosure in the Age of AI](https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1403904)

□ How managers' disclosure decisions cater to the rise of machine readers?

- Adopt machine-readable financial reporting formats that may have little impact on human readability
- Managers change their language styles to cater to large language models, such as Google BERT and FinBERT.
- Managers make their presentation styles (particularly voice features) more favorable to machine algorithm.

Human Readability vs. Machine

```
<TR VALIGN="BOTTOM">
<TH COLSPAN=4 ALIGN="LEFT" style="font-family:times;"><FONT SIZE=2><B>Discount Rate Sensitivity Analysis</B></FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=2>&nbsp;</FONT></TH>
<TH COLSPAN=5 ALIGN="CENTER" style="font-family:times;"><FONT SIZE=2><B>Table&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;1</B></FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=2>&nbsp;</FONT></TH>
</TR>
<TR style="font-size:1.5pt;" VALIGN="BOTTOM">
<TH COLSPAN=10 ALIGN="CENTER" style="font-family:times;border-bottom:solid #000000 1.0pt;">&nbsp;</TH>
<TH style="font-family:times;">&nbsp;</TH>
</TR>
<TR VALIGN="BOTTOM">
<TH ALIGN="LEFT" style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT></TH>
<TH COLSPAN=8 ALIGN="CENTER" style="font-family:times;border-bottom:solid #000000 1.0pt;"><FONT SIZE=1><B>Impact of </B></FONT></TH>
<TH style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT></TH>
</TR>
<TR VALIGN="BOTTOM">
<TH ALIGN="LEFT" style="font-family:times;"><FONT SIZE=1>(dollars in thousands)</FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT></TH>
<TH COLSPAN=2 ALIGN="RIGHT" style="font-family:times;"><FONT SIZE=1><B>Base<BR>
Discount<BR>
Rate</B></FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT></TH>
<TH COLSPAN=2 ALIGN="RIGHT" style="font-family:times;"><FONT SIZE=1><B>Discount<BR>
Rate<BR>
25 Basis<BR>
Point<BR>
Increase</B></FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT></TH>
<TH COLSPAN=2 ALIGN="RIGHT" style="font-family:times;"><FONT SIZE=1><B>Discount<BR>
Rate<BR>
25 Basis<BR>
Point<BR>
Decrease</B></FONT><BR></TH>
<TH style="font-family:times;"><FONT SIZE=1>&nbsp;</FONT></TH>
</TR>
<TR style="font-size:1.5pt;" VALIGN="BOTTOM">
<TH COLSPAN=10 ALIGN="CENTER" style="font-family:times;border-bottom:solid #000000 1.0pt;">&nbsp;</TH>
<TH style="font-family:times;">&nbsp;</TH>
</TR>
```

Machine
readability

High machine readability while maintaining human readability at the same level.

HTML as in a web browser (for the reader's convenience, the following picture shows the contents of the above scripts if shown as an HTML in a web browser⁵):

	Table 1		
	Base Discount Rate	Impact of 25 Basis Point Increase	Discount Rate 25 Basis Point Decrease
(dollars in thousands)			
2011 Net Periodic Benefit Cost	5.75%	\$ (20)	\$ 219
Benefit Plan Obligations as of December 31, 2011	5.04%	(3,214)	3,678
Estimated 2012 Net Periodic Benefit Cost	5.04%	(32)	16

Human
readability

Human Readability vs. Machine

Readability

Excerpt 1. APPLEBEES INTERNATIONAL INC, CIK: 0000853665, March 30, 2005

We opened 32 new company Applebee's restaurants in 2004 and anticipate opening at least 40 new company Applebee's restaurants in 2005, excluding up to eight restaurants that were closed in 2004 by a former franchisee which we may re-open in Memphis, Tennessee. The following table shows the areas where our company restaurants were located as of December 26, 2004:

Area	
New England (includes Maine, Massachusetts, New Hampshire, New York, Rhode Island and Vermont).....	65
Detroit/Southern Michigan.....	62
Minneapolis/St. Paul, Minnesota.....	58
St. Louis, Missouri/Illinois.....	47
North/Central Texas.....	45
Virginia.....	42
Kansas City, Missouri/Kansas.....	33
Washington, D.C. (Maryland, Virginia).....	29
San Diego/Southern California.....	20
Las Vegas/Reno, Nevada.....	15
Albuquerque, New Mexico.....	8

	424

Human
readability

(omitted)

```
<TYPE>EX-10
<SEQUENCE>4
<FILENAME>form10kexhf_032905.htm
<DESCRIPTION>EXHIBIT 10.2
<TEXT>
<HTML>
<HEAD>
<TITLE>Exhibit 10.2</TITLE>
```

Machine
readability

Low machine readability while maintaining human readability at the same level.

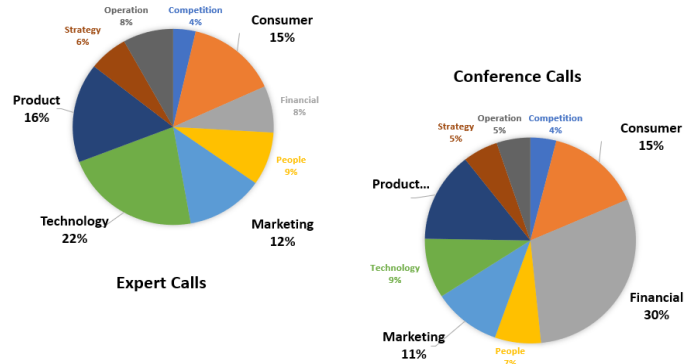
Source:

Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2023). How to talk when a machine is listening: Corporate disclosure in the age of AI. *The Review of Financial Studies*, 36(9), 3603-3642.

Stage II: Related AI research

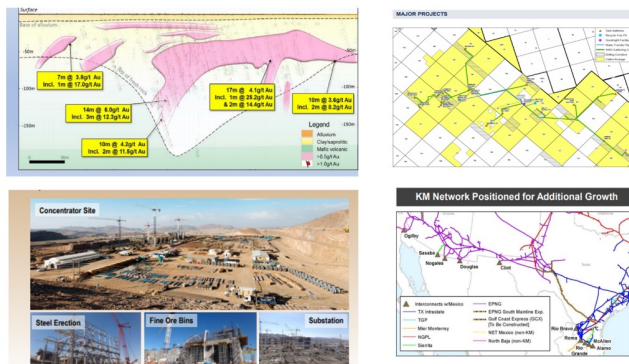
3. Expert Network Calls

- Topic Distribution of Expert Calls and Conference Calls

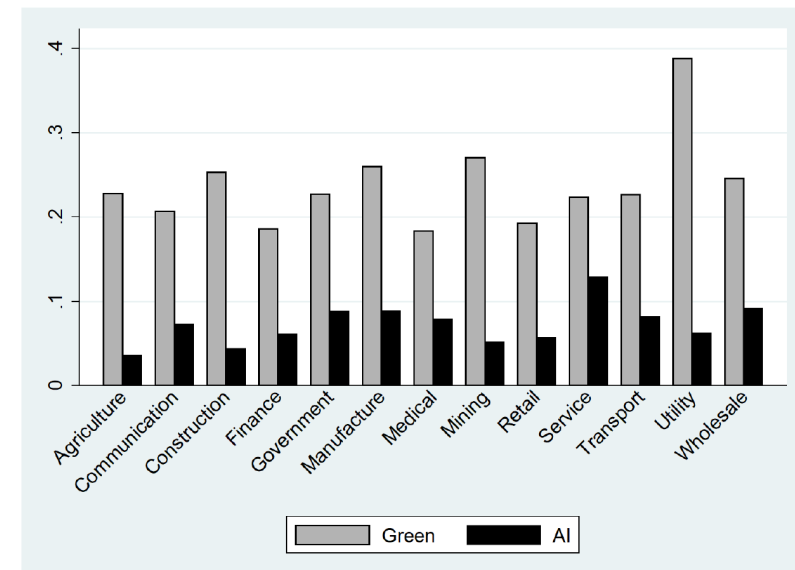


4. Image analysis

- Visual Information from presentation slides and AI divide



5. Executives learning about future AI and Climate investment



What is AI and big data in capital markets Research?

Theory work on blockchain

- AI and Fintech for decentralized finance
 - Industry: Metaverse, Blockchain, Web3, smart contracts, decentralized organization, Blockchain, P2P)
 - Academic research: more empirical work are coming
 - New Curriculums in business schools

What is AI and big data in capital markets Research?

■ AI and Alternative data for decision making

Relative mature area:

- AI and Alternative data for investment, stock returns and asset management .e.g., ML and NLP in finance
- AI and Alternative data for banks, loan contract and credit risk

Relative emerging area:

- **Public vs private firms**
 - AI and Alternative data for entrepreneurs or private firms/startups/small business
- **Institutional vs retail/household**
 - AI and Alternative data /emerging technologies for institutional investors vs household finance: AI access to retail investors, alternative lending, digital payments, robot advisor
- **AI for emerging societal challenges** such as sustainability and supply chain
 - AI and Alternative data for sustainable and social finance: **Climate Tech**, social equity and AI fairness
 - AI and Alternative data for supply chain stability
 - AI and Alternative data for cyber security

■ Policy for AI and Alternative data - RegTech: discrimination, gender issues, AI fairness

Classification of AI and Fintech Research



Type 1

Using Artificial Intelligence / Machine Learning as *Tools* to Calculate a New Variable

- The research question isn't centered on an AI topic
- AI/ML is used as a tool
- E.g. When predicting stock returns? ML vs OLS for prediction performance



Type 2

Research Questions Contain Artificial Intelligence / Machine Learning

- AI and ML are the central questions
- A growing research area
- E.g. How analysts use AI (Man+ Machine) to help them survive and thrive?

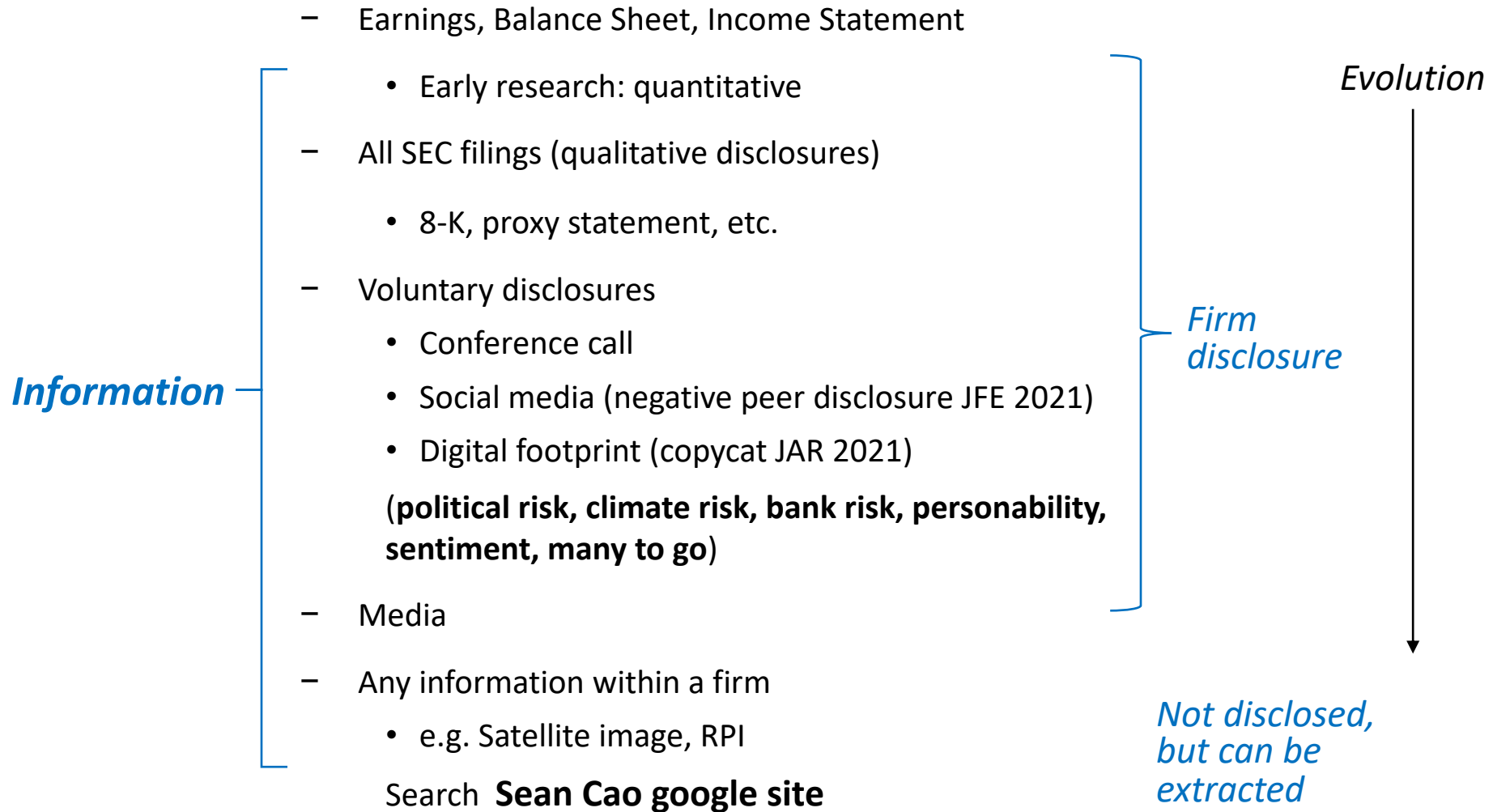


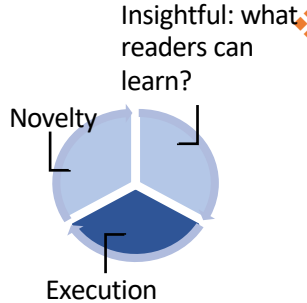
Type 3

Policy for AI

- State and firm-level policy pertaining to AI and big data
- No AI/ML tools are used
- The traditional diff-in-diff design is applied

Big data in Accounting and Finance vs CS scholars



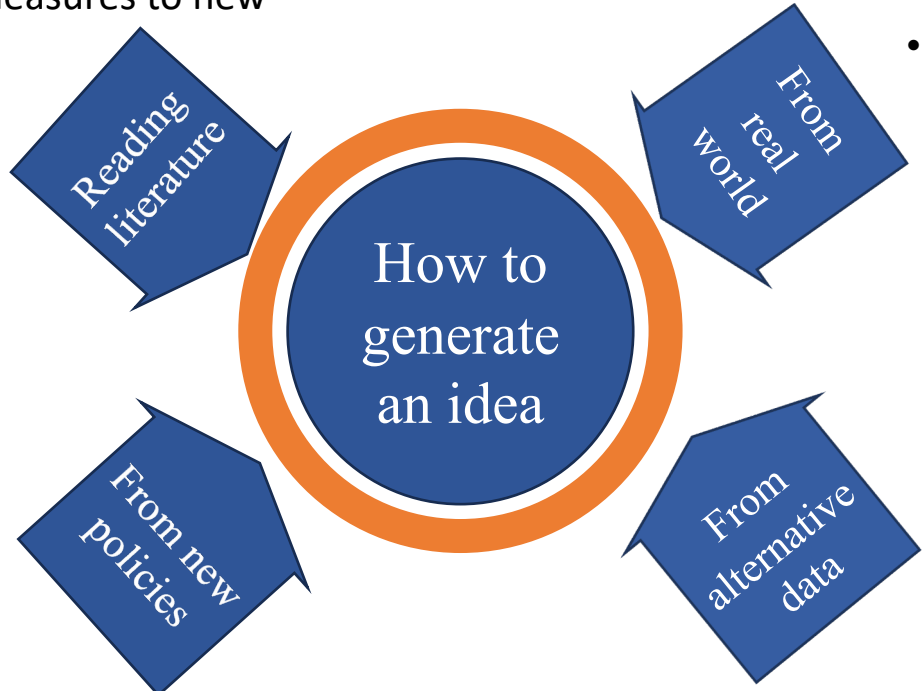
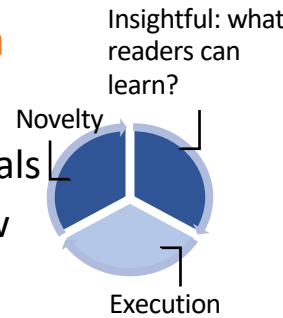


❖ Approach 1: Literature extension

- Find gaps from literature
- Offer new explanations
- Test existing theory in new settings
- Apply existing measures to new questions

❖ Approach 2: Real-world observation

- Read news
- Communicate with industry professionals
 - Identify a new problem/study a new phenomenon
 - Revisit old questions with improved methods (with caution for academic interests)

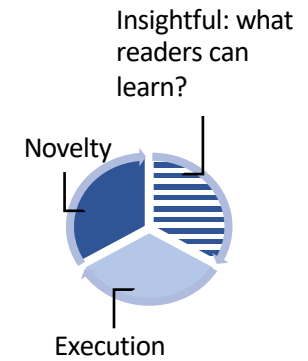
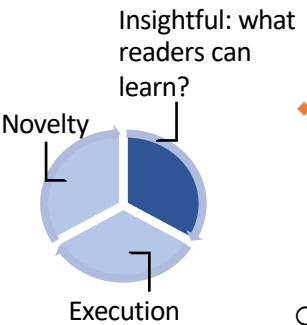


❖ Approach 4: New policies

- State Policy: potential competing papers
- Firm Policy
- High hurdle for empirical rigor in making policy implications

❖ Approach 3: Alternative data

- Novel data DOES NOT guarantee novel questions - thoughtful consideration is essential
- AI tool can help to implement the idea but not a selling point



Analytics for Finance and Accounting:
Data Structures and Applied AI



Sean Cao

University of Maryland

Wei Jiang

Emory University

Lijun Lei

*University of North Carolina
at Greensboro*

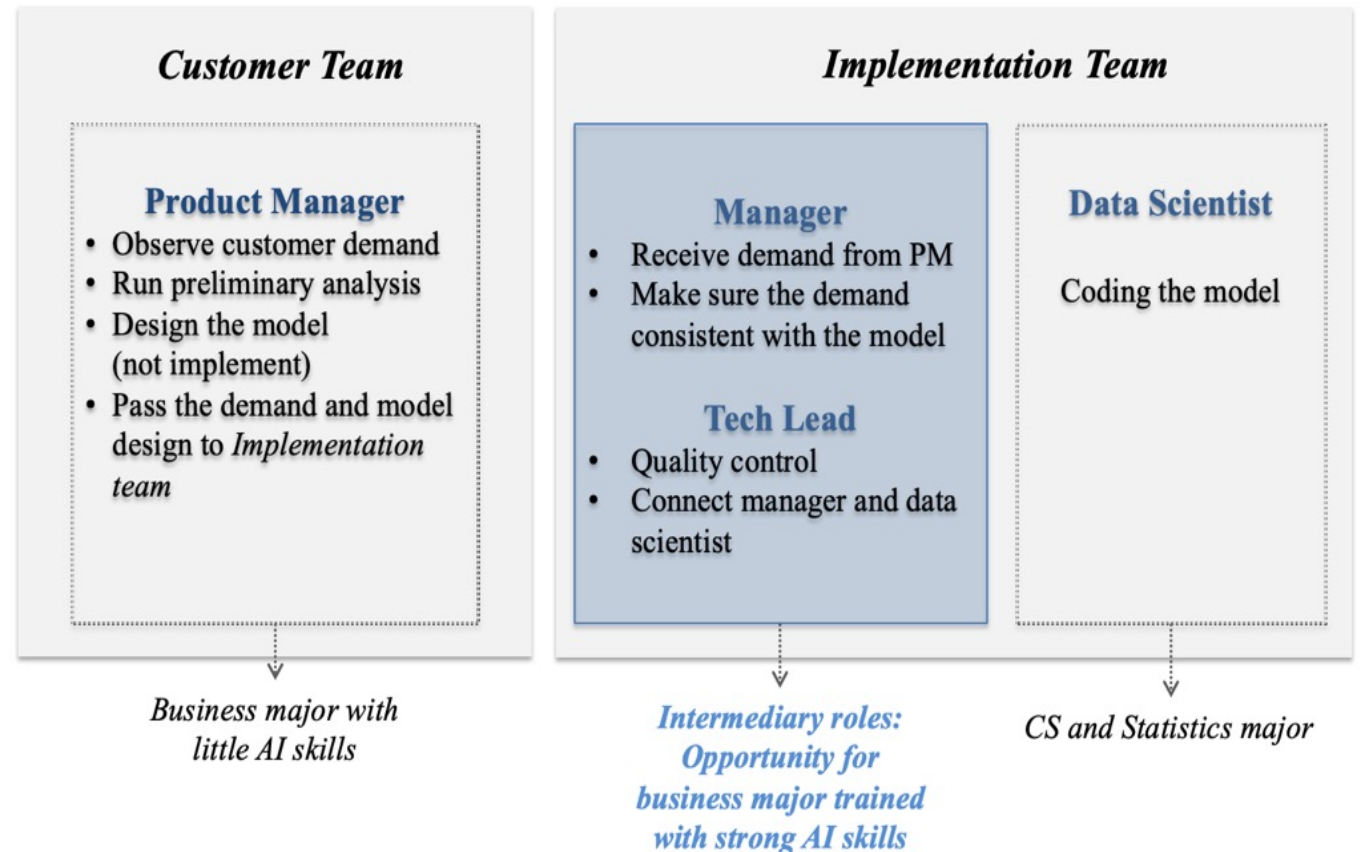
Analytics for Finance and Accounting:

Data Structures and Applied AI

Domain talent vs Data analytics talent

Training students with AI + Complexity of Business Domain Knowledge

- Data analytics in accounting and finance
- Timeline: it's URGENT.
- Over 50% of finance and accounting students may be replaced by AI (Cao et al., 2022, working paper).



Structure of the book

Part 1: Comprehending alternative financial data

Part 2: Familiarizing with emerging technologies

Part 3: Understanding the use cases for data and technology

Part 1: Comprehending domain data: where to get product/risk information of firms?

Apple Inc.'s 2021 10-K. It introduces the company background and provides information on Apple's main products.

Figure 1 Item 1 of Apple Inc. 2021 Form 10-K

PART I
Item 1. Business
Company Background
The Company designs, manufactures and markets smartphones, personal computers, tablets, wearables and accessories. The Company's fiscal year is the 52- or 53-week period that ends on the last Saturday of September. The Company was established in 1977.
Products
iPhone
iPhone® is the Company's line of smartphones based on its iOS operating system. During 2020, the Company released four new iPhone models with 5G technology; iPhone 12 and iPhone 12 Pro were available starting in September 2020, and iPhone 12 mini and iPhone 12 Pro Max are both expected to be available in November 2020.
Mac
Mac® is the Company's line of personal computers based on its macOS® operating system. During 2020, the Company released a fully redesigned Mac Pro®, and updated versions of its MacBook Air®, 13-inch MacBook Pro and 27-inch iMac®.
iPad
iPad® is the Company's line of multi-purpose tablets based on its iPadOS® operating system. During 2020, the Company released an eighth-generation iPad and introduced an all-new iPad Air®, which was available in September 2020.
Wearables, Home and Accessories
Wearables, Home and Accessories includes AirPods®, Apple TV®, Apple Watch®, Beats® products, HomePod®, iPod touch and other accessories. AirPods are the Company's wireless headphones that interact with Siri®. During 2020, the Company released the Company's line of smart watches based on its watchOS® operating system. In September 2020, the Company released the Apple Watch SE. In October 2020, the Company announced HomePod mini™, which is expected to be available in November 2020.

Figure 2. Item 1A of Apple Inc. 2021 Form 10-K

Item 1A. Risk Factors
The following discussion of risk factors contains forward-looking statements. These risk factors may be important to understanding other statements in this Form 10-K. The following information should be read in conjunction with Part II, Item 7, "Management's Discussion and Analysis of Financial Condition and Results of Operations" and the consolidated financial statements and accompanying notes in Part II, Item 8, "Financial Statements and Supplementary Data" of this Form 10-K.
The business, financial condition and operating results of the Company can be affected by a number of factors, whether currently known or unknown, including but not limited to those described below, any one or more of which could, directly or indirectly, cause the Company's actual financial condition and operating results to vary materially from past, or from anticipated future, financial condition and operating results. Any of these factors, in whole or in part, could materially and adversely affect the Company's business, financial condition, operating results and stock price.
Because of the following factors, as well as other factors affecting the Company's financial condition and operating results, past financial performance should not be considered to be a reliable indicator of future performance, and investors should not use historical trends to anticipate results or trends in future periods.
<i>The Company's business, results of operations, financial condition and stock price have been adversely affected and could in the future be materially adversely affected by the COVID-19 pandemic.</i>
COVID-19 has spread rapidly throughout the world, prompting governments and businesses to take unprecedented measures in response. Such measures have included restrictions on travel and business operations, temporary closures of businesses, and quarantines and shelter-in-place orders. The COVID-19 pandemic has significantly curtailed global economic activity and caused significant volatility and disruption in global financial markets.
The COVID-19 pandemic and the measures taken by many countries in response have adversely affected and could in the future materially adversely impact the Company's business, results of operations, financial condition and stock price. Following the initial outbreak of the virus, the Company experienced disruptions to its manufacturing, supply chain and logistical services provided by outsourcing partners, resulting in temporary iPhone supply shortages that affected sales worldwide. During the course of the pandemic, the Company's retail stores, as well as channel partner points of sale, have been temporarily closed at various times. In many cases, where stores and points of sale have reopened they are subject to operating restrictions to protect public health and the health and safety of employees and customers. The Company has at times required substantially all of its employees to work remotely.
The Company is continuing to monitor the situation and take appropriate actions in accordance with the recommendations and requirements of relevant authorities. The full extent of the impact of the COVID-19 pandemic on the Company's operational and financial performance is currently uncertain and will depend on many factors outside the Company's control, including, without limitation, the timing, extent, trajectory and duration of the pandemic, the development and availability of effective treatments and vaccines, the imposition of and compliance with protective public safety measures, and the impact of the pandemic on the global economy and demand for consumer products. Additional future impacts on the Company may include, but are not limited to, material adverse effects on: demand for the Company's products and services; the Company's supply chain and sales and distribution channels; the Company's ability to execute its strategic plans; and the Company's profitability and cost structure.
To the extent the COVID-19 pandemic adversely affects the Company's business, results of operations, financial condition and stock price, it may also have the effect of heightening many of the other risks described in this Part I, Item 1A of this Form 10-K.

Part 1: Comprehending domain data: where to sales contract or loan contracts of firms?

Figure 1. Earnings announcement tweet by Microsoft



Figure 1. Item 1.01 filed by Amazon Inc.

ITEM 1.01. ENTRY INTO A MATERIAL DEFINITIVE AGREEMENT.

On September 5, 2018, Amazon.com, Inc. (the "Company"), Bank of America, N.A. as administrative agent, and the lenders party thereto entered into a credit agreement (the "Credit Agreement"). The Credit Agreement provides the Company with an unsecured revolving credit facility with a borrowing capacity of up to \$2.0 billion. The term of the Credit Agreement is two years, but it may be extended for up to three additional one-year terms if approved by the lenders.

The initial interest rate applicable to outstanding balances under the Credit Agreement is the London interbank offered rate ("LIBOR") plus 0.625%, with a commitment fee of 0.06% on the undrawn portion of the credit facility under our current credit ratings. If the Company's credit ratings are downgraded these could increase to as much as LIBOR plus 1.00% and up to 0.10%, respectively.

Borrowings under the Credit Agreement may be used for working capital, capital expenditures, acquisitions, and other corporate purposes. The Company currently has no borrowings outstanding under the Credit Agreement, but expects to borrow under the Credit Agreement from time-to-time in the ordinary course of business.

The Credit Agreement contains customary representations and warranties, covenants, and events of default, but does not contain financial covenants. Upon an event of default that is not cured within applicable grace periods or waived, any unpaid amounts under the Credit Agreement may be declared immediately due and payable and the commitments may be terminated.

The foregoing description is qualified by reference to the full text of the Credit Agreement, which is filed as Exhibit 10.1 to this Current Report on Form 8-K.

The financial institutions party to the Credit Agreement and their respective affiliates are full service financial institutions engaged in various activities, which may include sales and trading, commercial and investment banking, advisory, investment management, investment research, principal investment, hedging, market making, brokerage, and other financial and non-financial activities and services. Certain of these financial institutions and their respective affiliates have provided, and may in the future provide, a variety of these services to the Company and to persons and entities with relationships with the Company, for which they received or will receive customary fees and expenses.

Figure 7. Reporting frameworks of Apple's 2021 ESG Report

About the report

Reporting year: This report focuses primarily on fiscal year 2020 activities, unless otherwise noted. All references to a year throughout the report refer to Apple's fiscal years, unless "calendar year" is specified. Apple's fiscal year is the 52- or 53-week period that ends on the last Saturday of September.

Alignment to reporting frameworks: The report leverages reporting frameworks and standards such as the Global Reporting Initiative (GRI), Sustainable Accounting Standards Board (SASB), and the Task Force on Climate-related Financial Disclosures (TCFD). Apple's ESG Index maps the disclosures in this report and other Apple publications against the metrics in these reporting frameworks, as relevant to our business.

Part 1: Comprehending domain data: Visual information beyond texts

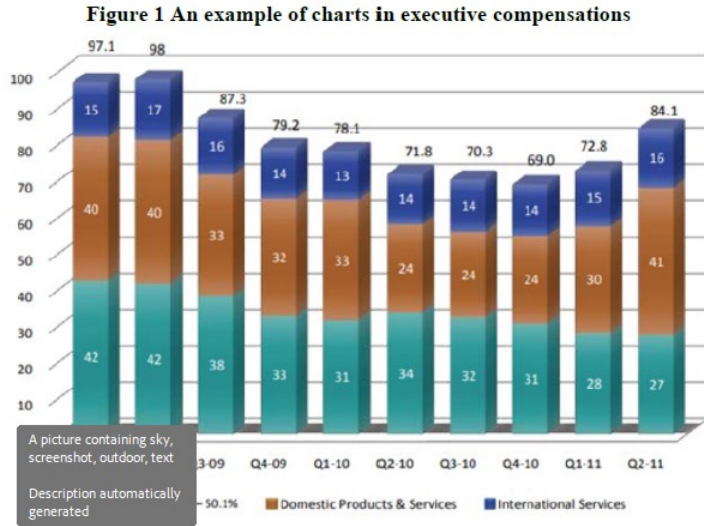


Figure 2 An example of images in executive presentations



Big data

Information

- Earnings, Balance Sheet, Income Statement
 - Early research: quantitative
- All SEC filings (qualitative disclosures)
 - 8-K, proxy statement, etc.
- Voluntary disclosures
 - Conference call
 - Social media (negative peer disclosure JFE 2021)
 - Digital footprint (copycat JAR 2021)

(political risk, climate risk, bank risk, personality, sentiment, many to go)
- Media
- Any information within a firm
 - e.g. Satellite image, RPI



Search **Sean Cao google site**

Evolution

Firm disclosure

Not disclosed, but can be extracted

Part 2: Emerging Technologies

3. Emerging AI technology in textual analysis	40
3.1. How to select and apply machine learning models.....	40
3.1.1 Data cleaning, parsing, and feature selection.....	40
3.1.2 Machine learning model selection	41
3.1.3 Hyperparameter tuning	43
3.1.4 Machine learning model performance evaluation.....	45
3.2. Pre-trained phrase level-word embedding	48
 Video Pre-trained phrase level analysis	
3.3. Pre-trained sentence level-word embedding.....	49
 Video Pre-trained sentence level analysis	
3.3.1 Bidirectional Encoder Representations from Transformers (BERT)	50
3.3.2 Generative pre-trained transformer (GPT).....	52
Appendix 3: Evaluating machine learning models	58

If you ask any questions for any firms...

For example, "What is the change in the compensation of Apple's CEO over the past three years?"

The screenshot shows a Google search interface. The search bar contains the question: "Question: What is the change in the compensation of Apple's CEO over the past three years?". Below the search bar are tabs for News, Perspectives, Images, Videos, Shopping, Books, Maps, Flights, and Finance. The search results show approximately 39,900,000 results in 0.50 seconds. The top result is from Fortune, titled "Apple's Tim Cook takes rare CEO pay cut after pushback. ...". Below the search results is a section titled "People also ask" with four related questions: "What is the annual compensation of the Apple CEO?", "What is the trend in CEO compensation?", "Did Apple CEO reduce his salary?", and "How much does the CEO of Apple make in 2023?". Below this is another search result from Investopedia titled "Apple CEO Tim Cook Receives a 40% Pay Cut".

Google search results



You

What is the change in the compensation of Apple's CEO over the past three years?

ChatGPT results:



ChatGPT

I'm sorry, but I don't have access to real-time data or information beyond my last knowledge update in January 2022. To find the most current information on the compensation of Apple's CEO over the past three years, I recommend checking Apple's latest annual reports or recent news articles, or visiting Apple's investor relations website. They typically release such information in their annual proxy statements or filings with the U.S. Securities and Exchange Commission (SEC).



You

Question: What is the change in the compensation of Apple's CEO over the past three years?



Alpha analyst's answers



Alpha Analysts

- In 2021, Tim Cook's total compensation was \$98,734,394. This included a \$3,000,000 salary, no bonus, \$82,347,835 in stock awards, \$12,000,000 as non-equity incentive plan compensation, and \$1,386,559 in all other compensation.
- In 2022, his total compensation decreased slightly to \$99,420,097. It consisted of a \$3,000,000 salary, no bonus, \$82,994,164 in stock awards, \$12,000,000 in non-equity incentive plan compensation, and \$1,425,933 in all other compensation.
- For 2023, Tim Cook's total compensation was reduced to \$63,209,845. His compensation breakdown included a \$3,000,000 salary, no bonus, \$46,970,283 in stock awards, \$10,713,450 as non-equity incentive plan compensation, and \$2,526,112 in all other compensation.

How to train LLM

- **Prompt Engineering:** (A student doesn't practice exams but take a note for the new exam) – parameters no change
 - **(Hard) Prompt-Tuning:** manually designed natural language, human-understandable prompts; Sensitive to choices of prompts
 - **(Soft) Prompt-Tuning**
 - **LM-BFF:** prompt-tuning + demonstration + template generation
 - **P-Tuning:** tuning the input (prompt) embeddings
Better performance via soft prompts
- **Model Fine-tuning** (A student takes many practice exams, become more skilled, and take the new exam) –parameter changes
 - **Prefix-Tuning:** only optimize the prefix embeddings
Improved training time and space efficiency, suitable for all models but requiring large training samples
 - **Finetune the whole model** – costly but yield the most optimal results.
- **Instruction Tuning:** human-guided reinforcement learning

“in-context” learning and fine-tuning

🟡 GPT “in-context” learning

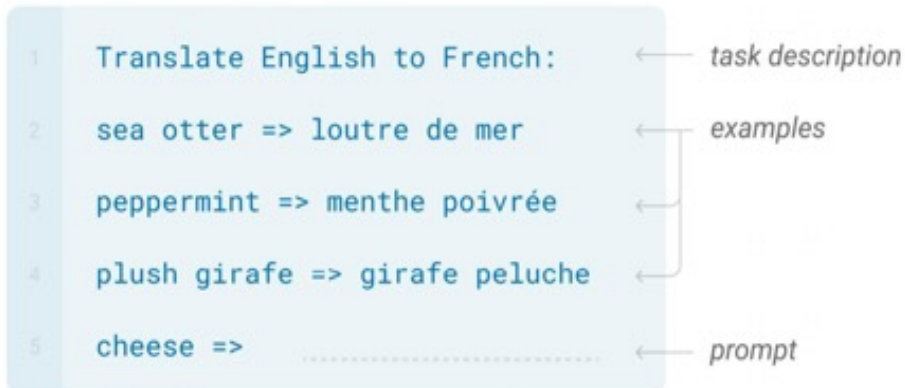
🟢 Zero-shot



🟢 One-shot

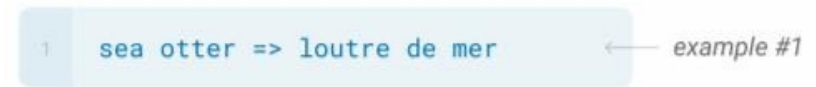


🟢 Few-shot



Parameters remain constant.

🟡 Traditional fine-tuning [Llama 2](#)



gradient update



gradient update



gradient update



Parameters are changed.

Prompt Tuning LLM

🕒 **Creating a specific LLM Account**

1. Create an account to access LLM API or directly get the open-source LLM.
2. Input a pair of prompts into the API:
 - Input: (a) Firm's conference call transcript
 - Input: (b) Questions
 - Output: Annotated answers (human identified climate sentence)

🕒 **Training the LLM Model**

The more examples (**input and output pairs**) provided, the better the LLM model will fit.

🕒 **Testing the LLM Model**

1. Provide a new input: a new conference call transcript and the same questions.
2. Check the response generated by the model.

Prompt Tuning – Prompt Example for Zero-Shot

```
[  
  {  
    "role": "system",  
    "content": "You are an AI analyst to help identify the climate risk related sentences from conference call transcript. Please return 1 if the input sentence is climate risk related and return 0 if the input sentence is not climate risk related. You will only provide the number 1 or 0 as the answer"  
  },  
  {  
    "role": "user",  
    "content": "Our commitment to reducing carbon emissions is a proactive measure to mitigate long-term climate risk"  
  },  
  {  
    "role": "assistant",  
    "content": ""  
  }  
]
```

Task Description

Input sentence to get the answer

Prompt Tuning – Prompt Example for One-Shot

```
[  
  {  
    "role": "system",  
    "content": "You are an AI analyst to help identify the climate risk related sentences from conference call transcript. Please return 1 if the input sentence is climate risk related and return 0 if the input sentence is not climate risk related. You will only provide the number 1 or 0 as the answer"  
  },  
  {  
    "role": "user",  
    "content": "We are integrating climate risk assessments into our strategic planning to ensure long-term sustainability and regulatory compliance."  
  },  
  {  
    "role": "assistant",  
    "content": "1"  
  },  
  {  
    "role": "user",  
    "content": "Our commitment to reducing carbon emissions is a proactive measure to mitigate long-term climate risk"  
  },  
  {  
    "role": "assistant",  
    "content": ""  
  },  
]
```

Task Description

One-shot Example

Input sentence to get the answer

Prompt Tuning – Prompt Example for Few-Shot

```
{  
  "role": "system",  
  "content": "You are an AI analyst to help identify the climate risk related sentences from conference call transcript. Please return 1 if the input sentence is climate risk related and return 0 if the input sentence is not climate risk related. You will only provide the number 1 or 0 as the answer"  
},  
{  
  "role": "user",  
  "content": "We are integrating climate risk assessments into our strategic planning to ensure long-term sustainability and regulatory compliance."  
},  
{  
  "role": "assistant",  
  "content": "1"  
},  
{  
  "role": "user",  
  "content": "Our quarterly earnings exceeded expectations, reflecting strong performance across all divisions."  
},  
{  
  "role": "assistant",  
  "content": "0"  
},  
{  
  "role": "user",  
  "content": "Our commitment to reducing carbon emissions is a proactive measure to mitigate long-term climate risk"  
},  
{  
  "role": "assistant",  
  "content": ""  
}
```

Task Description

Example 1

Example 2

Input sentence to get the answer

Few-shot Examples

How to train LLM

- **GPT-based or Open-source LLM:** (A student doesn't practice exams but take a note for the new exam) – parameters no change
 - **Prompt-Tuning:** manually designed natural language prompts
Human-understandable prompts; Sensitive to choices of prompts
(Soft) Prompt-Tuning
 - **LM-BFF:** prompt-tuning + demonstration + template generation
 - **P-Tuning:** tuning the input (prompt) embeddings
Better performance via soft prompts
- **Open-source LLM, such as Llama** (A student takes many practice exams, become more skilled, and take the new exam) –parameter changes
 - **Prefix-Tuning:** only optimize the prefix embeddings
Improved training time and space efficiency, suitable for all models but requiring large training samples
 - **Finetune the whole model** – costly but yield the most optimal results.
- **Instruction Tuning:** human-guided reinforcement learning

Videos available for each chapter and coding


See more at [Sean Cao google site](#)

Textbook for download

- [Analytics for Finance and Accounting: Data Structures and Applied AI](#)

Tutorial videos for download

- Each chapter of the textbook has tutorial videos for instructors and students. Please click the video link in the table of contents of the textbook.

4. Analyzing Conference Calls	59
4.1. Data in conference calls	59
 Video Data structure of conference calls	
4.2. Standard dependence parser.....	62
4.3. Empirical example: Contrasting conference calls and expert calls	65
Appendix 3: Using GPT for topic modeling.....	68

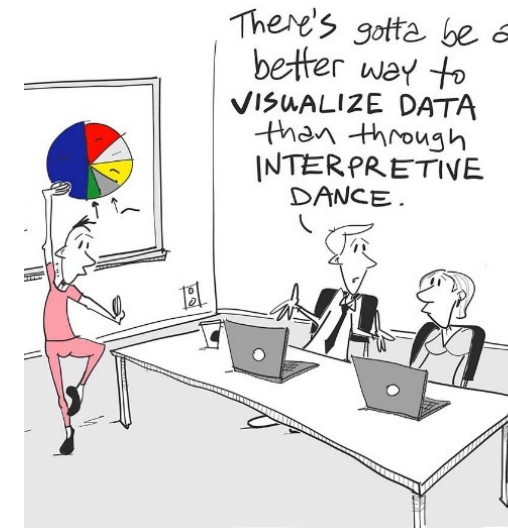
 [Video Applying GPT to analyze conference call transcripts using both API and web interface](#)

Slides for download

- [Tutorial slides](#)

Python tutorial for web crawling and fundamental textual analysis:

- [Crawling SEC Edgar](#)
- [Parsing 10-K filings](#)
- [Applying GPT: analyzing conference call transcripts using both API and web interface](#)



Analytics for Finance and Accounting:
Data Structures and Applied AI



Sean Cao
University of Maryland
Wei Jiang
Emory University
Lijun Lei
*University of North Carolina
at Greensboro*

How to use the textbook?

Free Textbook

- **For a new course:** an introductory course on AI literacy, specifically tailored for students in accounting, finance and other majors.
- **For traditional courses:** adopt one or two chapters as supplement
 - *E.g., Corporate Finance or Investment:* chapter 8k will be great supplement
 - *E.g., Intermediate Financial Accounting,* Chapter 10-K

Ongoing effort

- Train Big-data talents
 - For PhD, big data and AI short courses (e.g., Indiana University, Lancaster, UT Dallas)
 - For undergraduate, masters, executive education, New AI literacy course using the textbook
 - Don't require any computer-science background; cultivate their interests
- Knowledge transfer between academia and industry
 - They learn from our research, and we learn their demand from practice and institutional details
 - I observe high collaboration demand from industry in this emerging area

Investment Companies

Stock Selection Process

- **Step 1: Identifying Potential Stocks**

- Positive screening
- Use proprietary trading strategies to select a limited number of potential stocks
- Some companies already using AI

- **Step 2: Human Analysis and Red Flag Identification**

- Negative screening
- Human review selected stocks → potential red flags missed in Step 1
- Hypothetical team size: 20 people
- Very few companies are using AI

- **Step 3: Risk Control**

- Assess additional risk/portfolio volatility if adding the stock from Step 2 to portfolio
- Hypothetical team size: 10 people

- **Step 4: Trade Execution**

- Minimize price impact and transaction costs of executing the order
- When the amount of transaction is large with potential price impact



What can be the solutions for middle-size firms to use AI?

Transfer learning: standing on the *shoulders of giants*

- Utilize pre-existing AI models instead of building from scratch on your own
- It is costly to train pre-existing AI (pre-trained) like GPT using
 - supervised (human guidance)
 - self-supervised (using neighboring words as guidance)
 - unsupervised (machine clustering)
- Fine-tune models to meet your specific business demands



Example 1: ImageNet

Objective: identify the content of the image

- Complicated, because there are many items in the picture (e.g., concentrator site, steel erection, etc.)
- Transfer learning: all items can be well identified using existing models developed by ImageNet
- With the knowledge of common items, it is much easier to understand the image above
- Metaphor: You have all LEGO blocks and just figure out how to them together



What can be the solutions for firms to use AI?

Example 2: GPT

- GPT understands the context and meaning of much publicly available data, such as all books, Wikipedia, Twitter, etc.
- You can think of GPT as an expert who processes and absorbs knowledge from all these data
- Opportunity for business:
 - GPT is not specifically trained on financial data like 8-K or proxy statements, it has been trained on common financial data like annual reports and conference calls
 - GPT was trained on firms' private data
- If you want GPT to perform a firm-specific task, such as analyzing financial information of a stock
 - you don't need to start GPT training from infancy
 - rather, train it as if it were a well-informed person who has already assimilated all the data
 - It just needs less training on the domain data comparing to training from infancy.

You're not training an infancy, but rather training a super intelligent expert with your domain-specific data.

Investment Companies

Stock Selection Process – with AI

- **Step 2: AI-Assisted Red Flag Identification**

- Transfer learning: assuming you already has an expert who read all Wikipedia, Twitter, books, etc.
- Your domain data: the notes and report that previous analyst identified red flags and documented their logic and thinking process
- Your expert (pre-training model like GPT) can learn from these notes and identify red flags with less human intervention
- Team size: 1 people

- **Step 3: AI-Assisted Risk Control**

- Your domain data: the notes that your risk team have previously conducted risk tests and evaluations and documented calculation process
- Your expert AI (pre-training model like GPT) can learn from these notes and perform risk control with less human intervention
- Team size: 1 people

- ❖ **Benefits of AI adoption:**

- **Labor cost reduction:** decrease team size from 30 people to 2 people
- **Business expansion:** Increased business capacity from 100 million to 1 billion

What is the deal for other companies, such as manufacturing companies?

- **Identity potential customers (firm or individual costumers)**
 - Transfer learning: assuming you already has an expert, who read all Wikipedia, Twitter, books, etc.
 - Your domain data:
 - the notes that sales teams identify potential clients based on experience and market research and their reasoning and experience.
 - You have the information on which notes brought in new customers and which note didn't work
 - so that AI knows which strategy works and which strategy didn't. Both notes (success and failure) are important.
 - Your expert (pre-training model like GPT) can learn historical notes and data to identify client and engagement strategies.
 - Pro:
 - Sales team down from 50 to 5 people with *larger* customer base
 - Human sales person will not always work but AI sales will, because they learn from human failures and succusses



What is the deal for other companies, such as manufacturing companies?

- **AI helps with budgeting**

- Your domain data:
 - the notes used by our managers to allocate resources
 - you use the ex-post evaluation of the effectiveness of previous budgets,
 - so that AI knows which budget works and which budget didn't. Both notes (success and failure) are important.
- Your expert (pre-training model like GPT) can learn historical notes and data to do efficient budget that will work.
- Pro:
 - Human budget will not always work but AI budget will, because they learn from human failures and successes
 - Budgeting team can be reduced from 20 to 3 people with more effective budgeting decisions.



Key Benefits

- AI models trained on the company's unique data
- **Benefit 1: training on your own data and meet your specific business demand**
 - Each business has their own unique needs. General LLM may not address.
 - Fine-tuning a general LLM with your own data and specific task requirement will meet your specific business needs
- **Benefit 2: potentially protected privacy**
 - Uploading data to third-party LLM like GPT can compromise data privacy
 - If you have strict data privacy, use open-source LLM such as Llama 3
 - For non-open-source LLMs, you don't have access to their parameters, so you cannot build your own LLM.
 - Trade off between privacy and model performance

Investment Companies

Stock Selection Process

- **Step 1: Identifying Potential Stocks**
 - Positive screening
 - Use proprietary trading strategies to select a limited number of potential stocks
 - Some companies already using AI
- **Step 2: Human Analysis and Red Flag Identification**
 - Negative screening
 - Human review selected stocks → potential red flags missed in Step 1
 - Hypothetical team size: 20 people
 - Very few companies are using AI
- **Step 3: Risk Control**
 - Assess additional risk/portfolio volatility if adding the stock from Step 2 to portfolio
 - Hypothetical team size: 10 people
- **Step 4: Trade Execution**
 - Minimize price impact and transaction costs of executing the order
 - When the amount of transaction is large with potential price impact



What can be the solutions for middle-size firms to use AI?

Transfer learning: standing on the *shoulders of giants*

- Utilize pre-existing AI models instead of building from scratch on your own
- It is costly to train pre-existing AI (pre-trained) like GPT using
 - supervised (human guidance)
 - self-supervised (using neighboring words as guidance)
 - unsupervised (machine clustering)
- Fine-tune models to meet your specific business demands



Example 1: ImageNet

Objective: identify the content of the image

- Complicated, because there are many items in the picture (e.g., concentrator site, steel erection, etc.)
- Transfer learning: all items can be well identified using existing models developed by ImageNet
- With the knowledge of common items, it is much easier to understand the image above
- Metaphor: You have all LEGO blocks and just figure out how to them together



What can be the solutions for firms to use AI?

Example 2: GPT

- GPT understands the context and meaning of much publicly available data, such as all books, Wikipedia, Twitter, etc.
- You can think of GPT as an expert who processes and absorbs knowledge from all these data
- Opportunity for business:
 - GPT is not specifically trained on financial data like 8-K or proxy statements, it has been trained on common financial data like annual reports and conference calls
 - GPT was trained on firms' private data
- If you want GPT to perform a firm-specific task, such as analyzing financial information of a stock
 - you don't need to start GPT training from infancy
 - rather, train it as if it were a well-informed person who has already assimilated all the data
 - It just needs less training on the domain data comparing to training from infancy.

You're not training an infancy, but rather training a super intelligent expert with your domain-specific data.

Investment Companies

Stock Selection Process – with AI

- **Step 2: AI-Assisted Red Flag Identification**

- Transfer learning: assuming you already has an expert who read all Wikipedia, Twitter, books, etc.
- Your domain data: the notes and report that previous analyst identified red flags and documented their logic and thinking process
- Your expert (pre-training model like GPT) can learn from these notes and identify red flags with less human intervention
- Team size: 1 people

- **Step 3: AI-Assisted Risk Control**

- Your domain data: the notes that your risk team have previously conducted risk tests and evaluations and documented calculation process
- Your expert AI (pre-training model like GPT) can learn from these notes and perform risk control with less human intervention
- Team size: 1 people

- ❖ **Benefits of AI adoption:**

- **Labor cost reduction:** decrease team size from 30 people to 2 people
- **Business expansion:** Increased business capacity from 100 million to 1 billion

What is the deal for other companies, such as manufacturing companies?

- **Identity potential customers (firm or individual costumers)**
 - Transfer learning: assuming you already has an expert, who read all Wikipedia, Twitter, books, etc.
 - Your domain data:
 - the notes that sales teams identify potential clients based on experience and market research and their reasoning and experience.
 - You have the information on which notes brought in new customers and which note didn't work
 - so that AI knows which strategy works and which strategy didn't. Both notes (success and failure) are important.
 - Your expert (pre-training model like GPT) can learn historical notes and data to identify client and engagement strategies.
 - Pro:
 - Sales team down from 50 to 5 people with *larger* customer base
 - Human sales person will not always work but AI sales will, because they learn from human failures and succusses



What is the deal for other companies, such as manufacturing companies?

- **AI helps with budgeting**

- Your domain data:
 - the notes used by our managers to allocate resources
 - you use the ex-post evaluation of the effectiveness of previous budgets,
 - so that AI knows which budget works and which budget didn't. Both notes (success and failure) are important.
- Your expert (pre-training model like GPT) can learn historical notes and data to do efficient budget that will work.
- Pro:
 - Human budget will not always work but AI budget will, because they learn from human failures and successes
 - Budgeting team can be reduced from 20 to 3 people with more effective budgeting decisions.



Key Benefits

- AI models trained on the company's unique data
- **Benefit 1: training on your own data and meet your specific business demand**
 - Each business has their own unique needs. General LLM may not address.
 - Fine-tuning a general LLM with your own data and specific task requirement will meet your specific business needs
- **Benefit 2: potentially protected privacy**
 - Uploading data to third-party LLM like GPT can compromise data privacy
 - If you have strict data privacy, use open-source LLM such as Llama 3
 - For non-open-source LLMs, you don't have access to their parameters, so you cannot build your own LLM.
 - Trade off between privacy and model performance