NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | Nanyang Business School

Dean's Distinguished Speaker Series

Public Lectures

LLMs in Financial Analysis:

From Promise to Practice

7 May 2026, Thursday 6.30 pm – 9 pm SGT Hybrid Event at NBS Auditorium

SPEAKER

CHARLES C.Y. WANG
Tandon Family Professor of Business Administration
Harvard Business School
Harvard University

MODERATOR

KELVIN K. F. LAW
Associate Professor
Division of Accounting
Nanyang Business School
Nanyang Technological University

Charles C.Y. Wang

Tandon Family Professor of Business Administration
Harvard Business School
Harvard University

Scope of This Talk

A non-technical seminar for financial professionals.

Target Audience

- ✓ **Financial Professionals and Educators:** Professionals who live in 10-Ks, earnings calls, and spreadsheets.
- ✗ **NOT for "Quants":** This is not about high-frequency trading or statistical arbitrage.

The Big Questions

A) What does AI mean for *me*?

B) How can I use AI to do my job better?

C) What distortions should I know about?

D) Will I be replaced?

Large Language Models

- Neural network trained on a corpus that predicts the next token given a prompt.

$$f_{\theta_{\mathcal{I}}}(\mathit{Prompt}) = p_{\theta_{\mathcal{I}}}(Y \mid \mathit{Prompt}), \qquad Y \sim p_{\theta_{\mathcal{I}}}(\cdot \mid \mathit{Prompt}).$$

- **How do they work?**
  - Strictly: trillions of parameters applied to the prompt — not useful for understanding.
  - **Need a mental model** to understand what they mean for investors and markets.

Mental Model for LLMs

Imagine predicting the last word of a detective novel:

"And the murderer is _____."

(Analogy due to Ilya Sutskever, OpenAI)

- **To predict that word, you must understand:**
  - The plot, the characters, the evidence presented
  - The conventions of the genre
  - How this particular author structures their stories

- **Predicting text ≈ understanding text**
  - LLMs trained to predict sequences of words
  - In doing so, develop representations of meaning, context, and reasoning

Distinctive New Opportunities

- **Best *starting* "mental model" for these models might be human minds**
  - Basic neural network architecture inspired by animal brain
  - Many speculate frontier LLMs are/will soon become "thinking machines"

- **If so: new, distinctive opportunity**
  - Bring computational **speed, neutrality, breadth**
  - …to tasks that previously required human verbal reasoning and judgment
    (firm-specific analyses, qualitative evaluations, etc.)

⇒ Financial analysis: a useful arena to assess the implications of LLMs for white-collar work.

The Analyst's Challenge and the AI Testbed

Analyst’s job increasingly demanding.

- **Disclosure Bloat:** Filings are longer, more complex, and less standardized than ever before.
- **Information Overload:** Analysts must cull insights from text, video, audio, and endless data streams.
- Manual “needle in a haystack” search hard to scale; synthesis of information time consuming.

The Noise in GAAP & The Rise of the Machine

Key Concept in Financial Analysis: Core Earnings

> **Core Earnings:** Recurring earnings stemming from a company's central business activities.

Vs. Non-Core Earnings

Earnings stemming from ancillary activities or transitory shocks
(e.g., currency fluctuations, M&A costs, legal settlements).

The "Art" of Adjustment

> *"One of the first things you find is that there are GAAP earnings... and then there was 'real' earnings... I would strip out anything that wasn't recurring... A video game company's net income was boosted by selling a building? That's a one-time deal."*

— Confessions of a Former Analyst (VietEcon)

Why Judgment is Required

This process is not algorithmic. It requires context:

- **Business Model Matters:** Crypto profits might be core for Block, but non-core for Tesla.
- **Context Matters:** Is "Restructuring" truly a one-time event, or does this company restructure every year?
- **Disclosure Matters:** Is the expense buried in COGS and disclosed in footnotes, or explicitly labeled "Other Operating Expenses"?

Case Study 1: CBS Corp (2016)

The "Operating" Item that isn't “Core”

- **Reported Net Income:** $1.261 Billion.
- **The Adjustment:** A $211 Million Pension settlement charge (16.7% of Income).
- **The Issue:** This appeared in "Operating Expenses," depressing Operating Income. (**CapIQ & CompuStat considered part of core earnings!**)
- **Why it matters:** Just because it's “above the line” (Operating Income) doesn't mean it's Core. This was a one-time transfer of liabilities, not a recurring cost of content production.

Consolidated Statements of Operations - USD ($) shares in Millions	12 Months Ended		
	Dec. 31, 2016	Dec. 31, 2015	Dec. 31, 2014
Income Statement [Abstract]			
Revenues	$ 13,166,000,000	$ 12,671,000,000	$ 12,519,000,000
Costs and Expenses [Abstract]			
Operating	7,956,000,000	7,911,000,000	7,689,000,000
Selling, general and administrative	2,124,000,000	1,961,000,000	1,971,000,000
Depreciation and amortization	225,000,000	235,000,000	250,000,000
Pension settlement charge (Note 15)	211,000,000	0	0
Restructuring and merger and acquisition-related costs (Note 5)	38,000,000	45,000,000	19,000,000
Other operating items, net	(9,000,000)	(139,000,000)	0
Total costs and expenses	10,545,000,000	10,013,000,000	9,929,000,000
Operating income	2,621,000,000	2,658,000,000	2,590,000,000
Interest expense	(411,000,000)	(392,000,000)	(363,000,000)
Interest income	32,000,000	24,000,000	13,000,000
Loss on early extinguishment of debt	0	0	(352,000,000)
Other items, net	(12,000,000)	(26,000,000)	(30,000,000)
Earnings from continuing operations before income taxes and equity in loss of investee companies	2,230,000,000	2,264,000,000	1,858,000,000
Provision for income taxes	(628,000,000)	(676,000,000)	(659,000,000)
Equity in loss of investee companies, net of tax	(50,000,000)	(34,000,000)	(48,000,000)
Net earnings from continuing operations	1,552,000,000	1,554,000,000	1,151,000,000
Net earnings (loss) from discontinued operations, net of tax (Note 4)	(291,000,000)	(141,000,000)	1,808,000,000
Net earnings	$ 1,261,000,000	$ 1,413,000,000	$ 2,959,000,000

Case Study 2: UPS (2012)

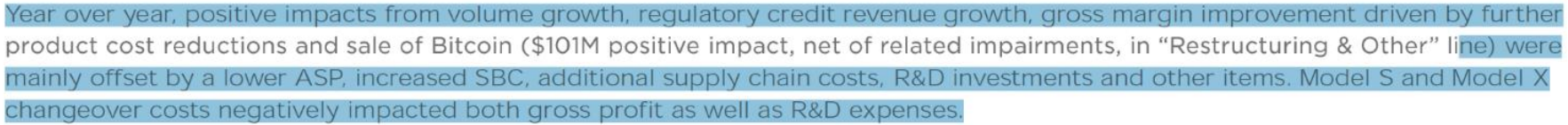
Hidden in the Aggregation

- **The Event:** A massive mark-to-market pension charge ($4.831 Billion) + 900M pension plan withdrawal charge
- **The Hiding Spot:** Disclosed in the **MD&A** (page 34 of 113) and footnotes. Aggregated into “Compensation and Benefits on the I/S. (**CapIQ & CompuStat considered part of core earnings!**)
- **The Lesson:** Don’t look at the face of the financial statements alone. Must read the text!

Case Study 3: Tesla (2021)

The Unsearchable Bitcoin Gain

- **Context:** Tesla buys Bitcoin (late 2020). Coin appreciates 85% in Q1 2021.
- **Event:** Record Q1 earnings (April 2021). Analysts ask: **"Is this just Bitcoin sales?"**
- **The Hunt:** No mention in the shareholder deck! "Ctrl+F 'Bitcoin'" → 0 results relating to Bitcoin profits!
- **The Reality:** A $101 Million profit was booked.


Year over year, positive impacts from volume growth, regulatory credit revenue growth, gross margin improvement driven by further product cost reductions and sale of Bitcoin ($101M positive impact, net of related impairments, in "Restructuring & Other" line) were mainly offset by a lower ASP, increased SBC, additional supply chain costs, R&D investments and other items. Model S and Model X changeover costs negatively impacted both gross profit as well as R&D expenses.

- **The Trick:** The disclosure was embedded as a **PICTURE** in the deck to make it text-unsearchable.

Stylized Facts: Rouen, So, & Wang (2021)

Understanding the Magnitude of the Problem

Stylized Fact #1: Frequency & Magnitude

Non-core earnings items are pervasive.

- **99% of firms** disclose at least one non-core earnings item over their history.
- **Frequency:** The average 10-K contains >6 non-core adjustments.
- **Trend:** This frequency has increased by 33% (6→8 items) from 1998 to 2017.

Economic Magnitude

The average total value of these adjustments amounts to approximately **19% of Net Income**.

Stylized Fact #2: The "Hidden" Iceberg

A substantial portion of these items are **not** on the Income Statement.

On the Face (Visible)

~50%

Items explicitly labeled on the Income Statement (e.g., “Restructuring Charge”).

Off the Face (Hidden)

~50%

Items buried in **Footnotes**, **MD&A**, or **Cash Flow Statements**.

(e.g., The UPS Pension Charge)

Trend: Growth in non-core earnings disclosures driven largely by items on the face of I/S (56%), but off-I/S items also grew (15%), mostly from MD&A disclosures.

Magnitude: Off-I/S adjustments commonly exceed 35% of net income!

The "Robo-Analyst" Approach

Increasing disclosure bloat + cognitive demand on analysts → opportunities to deploy AI

New Constructs (NC): A pioneer in using technology to scale fundamental analysis.

- **Hybrid Model:** Combined human judgment with systematic machine extraction.
- **Method:** Machines parse 10-Ks; humans verify complex edge cases.
- **Goal:** Systematic removal of accounting distortions to find Economic Earnings.

Ticker	Name	Type	Sector / Style	Mkt Val / AUM[1]	Div Yield	Equity Rating		ROIC	FCF Yield	Price to EBV	GAP	Credit Rating		Analyst Notes[2]
XLV	State Street Select Sector SPDR Trust: Health Care Select Sector SPDR Fund	ETF	Healthcare	$31,994	1.41%	Very Attractive		19%	5%	1.6	24	Not Available	-	Nov 26, 2021
AAPL	Apple Inc.	Stock	Technology	$2,680,712	0.5%	Attractive		237%	3%	1.8	21	Very Attractive		
NFLX	Netflix Inc.	Stock	Technology	$273,067	-	Neutral		20%	-0%	3.0	29	Attractive		Nov 23, 2021
RIVN	Rivian Automotive Inc	Stock	Consumer Cyclicals	$95,714	-	Unattractive Suspended		-30%	-2%	-4.2	> 100	Neutral Suspended		Nov 23, 2021
PLAXX	PACE Select Advisors Trust: PACE Large Co Growth Equity Investments	Mutual Fund	Large Cap Growth	$1,278	-	Very Unattractive		33%	1%	3.3	40	Not Available	-	

Learnings

- Robo-analyst tech created superior Core Earnings metrics at scale with speed
  - Analysts don’t process this information in their earnings forecasts on a timely basis
  - Particularly the adjustments hidden in the footnotes
  - Markets slow to impound core earnings information into prices, particularly when hidden in footnotes

- But, extraction remained expensive (still required significant human labor):
  access to platform costs tens of thousands.

Part 1: The Promise

Democratizing Analysis

$20~200 / month

vs. >$20,000 / year

ChatGPT5	Gemini-3, Claude Opus-4.7
400k	1m
Input Tokens	Input Tokens
~600-800 pages of text	*~1200-1500 pages of text*

Sufficient for financial analysis!!!

Could generative AI (GPT-5, Claude Opus 4.7, Gemini-3) offers a general-purpose solution that drastically lowers information processing costs for financial analysis? White collar work?

What an LLM Does That the Robo-Analyst Couldn't

Three capability shifts make general-purpose models useful where rules engines couldn't reach.

1

No rules engine

Reads any 10-K cold — no taxonomy, no hand-coded extractors, no per-firm tuning.

2

Multi-modal

Text, tables, images, and charts processed jointly. Disclosures hidden in pictures stop being invisible.

3

Explains itself

Returns the answer with natural-language reasoning.

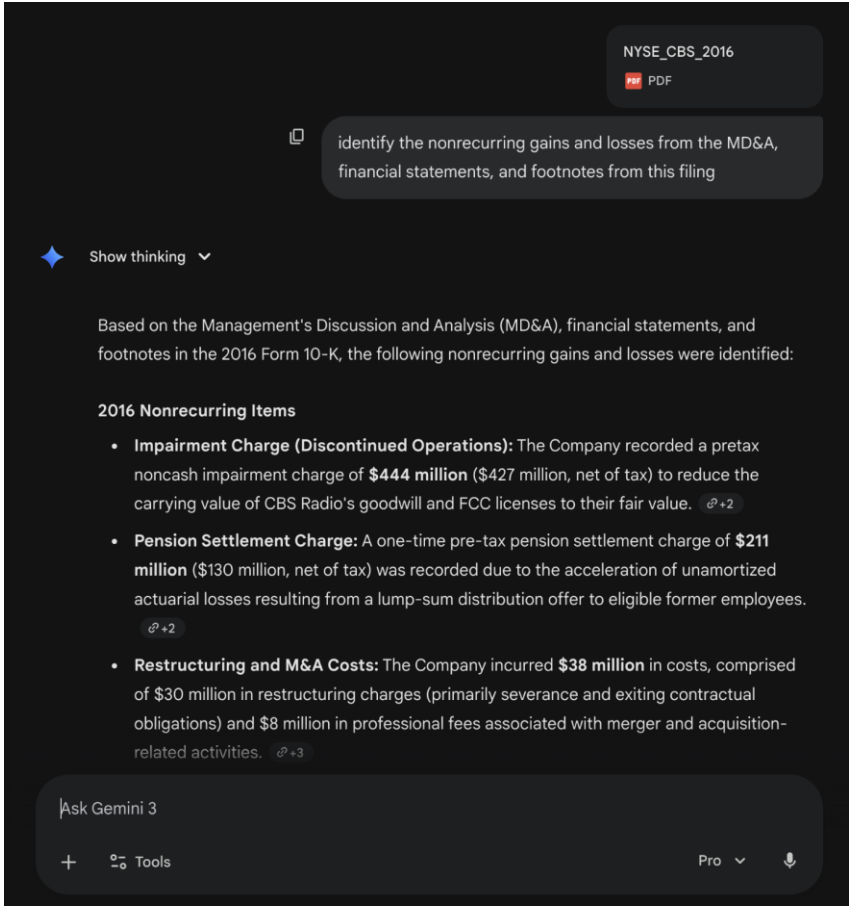
Citadel CFO Umesh Subramanian @ Milken Institute, 2025, on how AI is being used:

“An intelligent assistant helping portfolio managers sift through torrents of news, filings, and research so they can get the insights that matter, faster.”

Applications: CBS Corp (2016)

Consolidated Statements of Operations - USD ($) shares in Millions	12 Months Ended		
	Dec. 31, 2016	Dec. 31, 2015	Dec. 31, 2014
Income Statement [Abstract]			
Revenues	$ 13,166,000,000	$ 12,671,000,000	$ 12,519,000,000
Costs and Expenses [Abstract]			
Operating	7,956,000,000	7,911,000,000	7,689,000,000
Selling, general and administrative	2,124,000,000	1,961,000,000	1,971,000,000
Depreciation and amortization	225,000,000	235,000,000	250,000,000
Pension settlement charge (Note 15)	211,000,000	0	0
Restructuring and merger and acquisition-related costs (Note 5)	38,000,000	45,000,000	19,000,000
Other operating items, net	(9,000,000)	(139,000,000)	0
Total costs and expenses	10,545,000,000	10,013,000,000	9,929,000,000
Operating income	2,621,000,000	2,658,000,000	2,590,000,000
Interest expense	(411,000,000)	(392,000,000)	(363,000,000)
Interest income	32,000,000	24,000,000	13,000,000
Loss on early extinguishment of debt	0	0	(352,000,000)
Other items, net	(12,000,000)	(26,000,000)	(30,000,000)
Earnings from continuing operations before income taxes and equity in loss of investee companies	2,230,000,000	2,264,000,000	1,858,000,000
Provision for income taxes	(628,000,000)	(676,000,000)	(659,000,000)
Equity in loss of investee companies, net of tax	(50,000,000)	(34,000,000)	(48,000,000)
Net earnings from continuing operations	1,552,000,000	1,554,000,000	1,151,000,000
Net earnings (loss) from discontinued operations, net of tax (Note 4)	(291,000,000)	(141,000,000)	1,808,000,000
Net earnings	$ 1,261,000,000	$ 1,413,000,000	$ 2,959,000,000

- **The Adjustment:** A $211 Million Pension settlement charge (16.7% of Income).
- **The Issue:** This appeared in "Operating Expenses," depressing Operating Income. (**CapIQ & CompuStat considered part of core earnings!**)

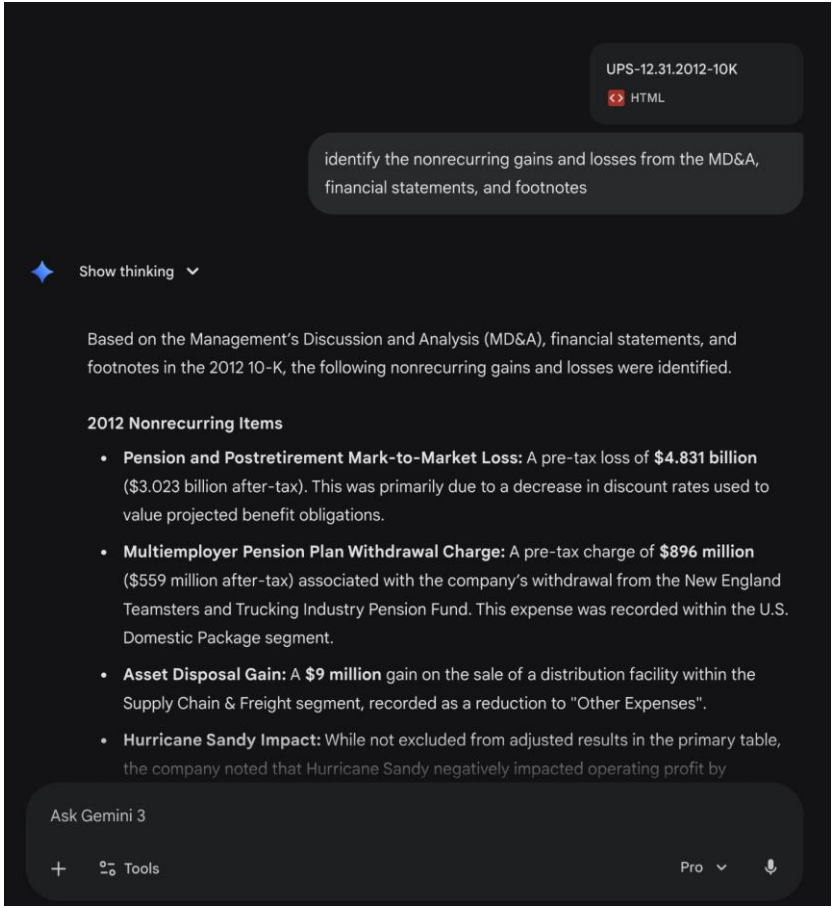


Applications: UPS (2012)

Operating Expenses:	
Defined Benefit Plans Mark-to-Market Charge	$ 4,831

	Years 2012
Revenue	$ 54,127
Operating Expenses:	
Compensation and benefits	33,102
Repairs and maintenance	1,228
Depreciation and amortization	1,858
Purchased transportation	7,354
Fuel	4,090
Other occupancy	902
Other expenses	4,250
Total Operating Expenses	52,784
Operating Profit	1,343
Other Income and (Expense):	
Investment income	24
Interest expense	(393)
Total Other Income and (Expense)	(369)
Income Before Income Taxes	974
Income Tax Expense	167
Net Income	$ 807

- **The Event:** A massive mark-to-market pension charge ($4.831 Billion) + 900M pension plan withdrawal charge
- **The Issue:** Hidden in the aggregation.

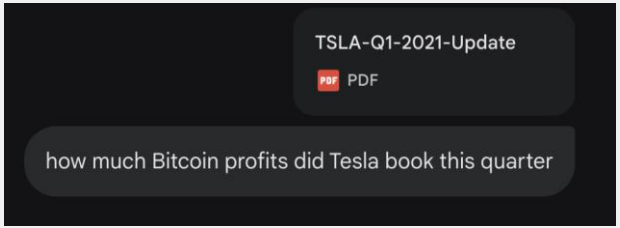


Applications: Tesla’s Bitcoin Gains (2021)

Tesla’s "Picture" Trick

Year over year, positive impacts from volume growth, regulatory credit revenue growth, gross margin improvement driven by further product cost reductions and sale of Bitcoin ($101M positive impact, net of related impairments, in “Restructuring & Other” line) were mainly offset by a lower ASP, increased SBC, additional supply chain costs, R&D investments and other items. Model S and Model X changeover costs negatively impacted both gross profit as well as R&D expenses.

LLMs’ Multimodal Power



Based on the document provided, Tesla booked a **$101 million** positive impact from the sale of Bitcoin in Q1 2021.

This gain was recorded in the "Restructuring & Other" line item of their financial summary and is calculated net of related impairments.

Part 2: The Pitfalls

PART 2 · PITFALL 1 OF 3 — Semantic Distortion and Overreach

Evidence from Shaffer & Wang (2026)

Opening the Black Box

Using LLMs as Core Earnings Analyst

Shaffer & Wang (2026)

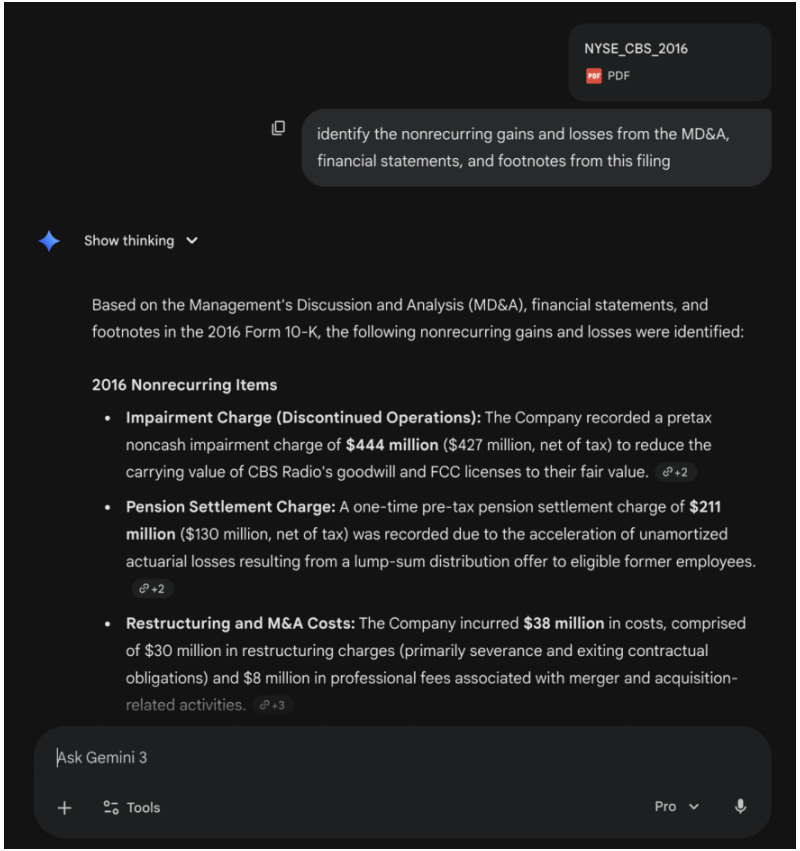
Applying LLMs to scale Core Earnings estimation.

The Experiment

Can a general-purpose LLM replicate the work of specialized financial analysts?

We tested two prompting strategies on thousands of 10-Ks:

1. **Lazy Analyst (One-Shot Prompt):** "Here is a 10-K. Find Core Earnings."
2. **Sequential Prompt:** Step-by-step guidance. (Impose reasoning path.)



The Prompts

Strategy A: Lazy Analyst

```
You are a financial analyst tasked with determining a company's
core earnings based on its 10-K filing. Core earnings represent the
persistent profitability of the company's central and ongoing
activities, exclusive of ancillary items and one-time shocks. This
concept aims to capture the owner's earnings - the sustainable,
recurring profitability that accrues to equity holders.

Please analyze the provided 10-K text and estimate the company's
core earnings. Start with the reported GAAP net income and make
adjustments you deem necessary based on the information in the 10-K.
Provide a clear explanation of your reasoning for each adjustment.
Additionally, to make it possible to extract your answer later,
please include the following tag at the end of your response, after
you finish your reasoning and calculation: "*Core Earnings
Calculation (final) = $[your determination]" where
[your determination] is the final core earnings amount you
calculate.

Here is the full 10-K text: [Cleaned 10-K Text Inserted Here]
```

Strategy B: Sequential Prompt

1. **Call 1: Identification of Non-Recurring Losses/Expenses**:

   You are an expert financial analyst with extensive experience.

   Here's a 10-K. Are there any nonrecurring/unusual expenses in the income statement, cash flow statement, footnotes, or MD&A? Be comprehensive and check your work twice.

   [Cleaned 10-K Text Inserted Here]

2. **Call 2: Identification of Non-Recurring Gains/Income**:

   Are there any nonrecurring/unusual income in the income statement, cash flow statement, footnotes, or MD&A? Be comprehensive and check your work twice.

3. **Call 3: Computation of Adjusted Earnings**:

   Based on the above, compute a new earnings measure. Start with net income, add back nonrecurring/unusual expenses, subtract nonrecurring/unusual income. No hypothetical values. Express in $ Millions.

   Provide a summary table like this example:

Example: Lawson Products, FY 2017

GPT-4o, given a precise definition of core earnings and the full 10-K:

- **Starting point:** GAAP Net Income $29.7M
- ✓ Gain on sale of property – $5.4M *(legitimate, R5)*
- ✓ Acquisition-related expenses + $0.7M *(legitimate, R4)*
- ✓ Tax valuation-allowance release – $21.2M *(legitimate, R5)*
- ✗ Interest expense + $0.6M
- ✗ D&A + $6.8M
- ✗ Stock comp + $3.1M

⇒ The model drifts toward enterprise/cash adjustments — not core earnings.

Lazy Analyst's Habits

Lawson isn't a one-off — across ~2,000 firms and 23 years, the same pattern emerges:

- Adjusts items (D&A, SBC, interest expense) it shouldn't; conflates enterprise/CF concepts.

One-Shot Definition (Lazy Analyst)	Sequential Summarization
1. Tax One-Time (R5)	1. Asset Sale G/L (R5)
2. Restructuring (R5)	2. Impairment (R4)
3. Impairment (R4)	3. Restructuring (R5)
4. Asset Sale G/L (R5)	4. Tax One-Time (R5)
5. M&A Costs (R4)	5. M&A Costs (R4)
6. Stock Compensation (R1) ⚠	6. Litigation (R4)
7. Depreciation (R1) ⚠	7. Debt Extinguishment (R4)
8. Interest Expense (R1) ⚠	8. Unrealized Fair Value (R4)
9. Litigation (R4)	9. FX G/L (R4)
10. Debt Extinguishment (R4)	10. Realized Investment G/L (R2)

⚠ *= Three R1 (low-quality, recurring) items in the top 10 — Sequential Summarization has none.*

Understanding the Sources of Failure

Two hypotheses for why the Seq Summ outperforms Lazy Analyst — and what we found:

- **Hypothesis A — Architectural:** autoregressive generation can't change course once initiated
  - Test: rerun the same prompt with o3 reasoning model (hidden chain-of-thought)
  - **Result: only partial improvement** ("bad adjustments" share: 20.5% → 15.8%)
- **Hypothesis B — Semantic:** "core earnings" triggers training-corpus associations (mgmt non-GAAP, EBITDA, Street earnings)
  - Test: same prompt, but use "recurring/unusual" instead of "core"
  - **Result: substantial improvement** ("bad adjustments" share: 20.5% → 8.3%)

⇒ Failure turns more on semantic distortion than architecture.

A Specific Reasoning Path

Sidestepping the polysemous term gets us most of the way — but not all the way:

- Lazy Analyst → Reasoning model: "bad adjustments" share **20.5% → 15.8%** *(architecture)*
- + "recurring/unusual" semantics: **20.5% → 8.3%** *(language)*
- **+ Sequential Summarization: 20.5% → 3.6%**

The residual gain — from **8.3% to 3.6%** — we attribute to imposing a **specific reasoning path**:

- Tell the model *how to proceed*, not *what we're going for conceptually*.
- Decompose the task: separate losses from gains, identification from aggregation.

What's Left? "Reaching" Behavior

Even after the harnessed workflow, "bad adjustments" rate isn't zero. Why?

- **Not** long-context fatigue: filing length is uncorrelated with "bad adjustments" rate.
- **The model *reaches*** — populates the table even when there's little to find.
  - Firms with **few real special items**: "bad adjustments" share **4.25%**
  - Firms with **many real special items**: "bad adjustments" share **1.80%**
    (2.4× difference; pattern repeats with firm size and high-quality adjustment count)
  - Consistent with instruction tuning (RLHF): models rewarded for being a "helpful assistant"

⇒ Assistant-tuning pressures the model to "do something" — even when restraint is warranted.

PART 2 · PITFALL 2 OF 3 — ARCHITECTURAL LIMITS

Evidence from Levy (2026)

Opening the Black Box

What LLMs are GOOD At

- **Needle-in-a-Haystack:** Locating and tabulating specific information (e.g., numbers, facts) in lengthy documents.
- **Representation Learning:** Extracting features (sentiment, tone, vocal delivery quality) from unstructured data in a data driven, objective manner.
- **Multimodal Analysis:** Unifying text, vision, and audio into a common representation space to "see" what text scrapers miss.

Levy's Insight

AI models excel at extracting fundamental latent characteristics from multi-modal content in a data-driven manner, enabling unified constructs across domains to deepen understanding of stakeholder perceptions (e.g., how text sentiment aligns with vocal cues or facial traits).

Numerical Reasoning (Balancing)

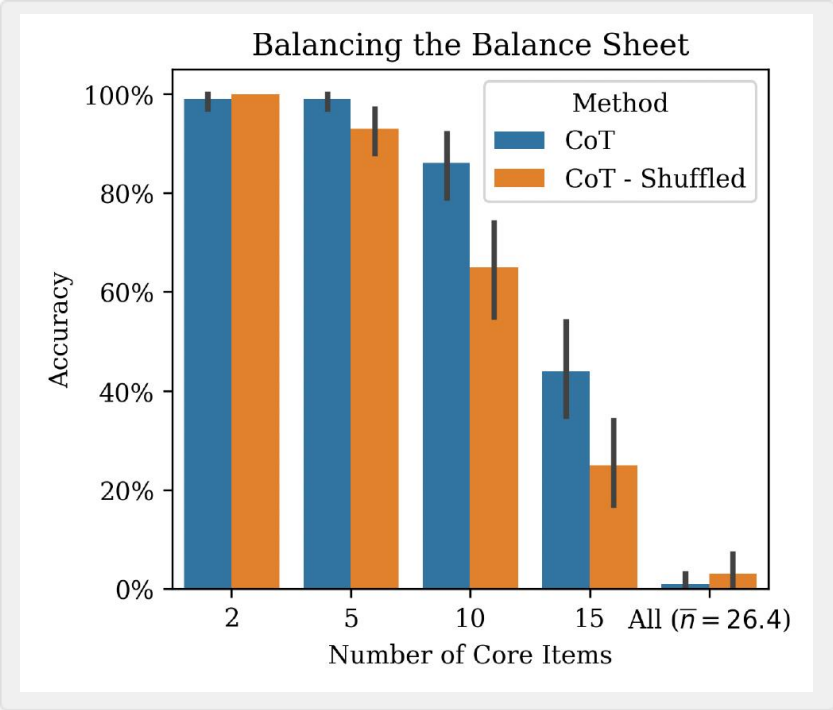
Can LLMs actually do arithmetic? **No.** They predict the next token.

Balance Sheet Experiment

Task: Tally numbers from a Balance Sheet.

- **Adding 2 Numbers:** High Accuracy (~99%).
- **Adding 10 Numbers:** Accuracy drops to ~20%.
- **Adding All Items (26+):** Accuracy drops to ~1-12%.

Implication: Don't trust LLMs to do computational work! Use a tool (python, calculator, abacus)!



Limitation 2: Look-Ahead Bias

Is the AI predicting earnings, or remembering them?

Levy's Digit Experiment

Asked GPT-4 to predict future earnings (increase or decrease) based on historical financials.

- **Original Data:** 60% Accuracy (Predictive).
- **Randomized Pennies:** When the *least significant digit* was randomized (e.g., $10.01 -> $10.04), accuracy dropped to **50% (Random Guess)**.

Conclusion: In this specific forecasting task, the model wasn't analyzing the trend; it recognized the specific number string from its training data.

THE THROUGH-LINE

Every failure we've seen is the same failure: the most likely next token wins.

Same mechanism in three axes.

- **CONCEPT** ✓ covered
  “Core earnings” becomes whatever the corpus thinks it means.
  Shaffer & Wang
- **NUMBERS** ✓ covered
  Arithmetic becomes the most likely next token.
  Levy
- **TIME** ✓ covered
  Historical forecasts become memorized digit strings.
  Levy
- **SPACE** ▸ up next
  What happens when the corpus is geographically lopsided?
  Cao, Wang & Yi — next

PART 2 · PITFALL 3 OF 3 — SPATIAL INFORMATION GAPS

Evidence from Cao, Wang, Yi (2026)

Foreign Bias in Financial Predictions

The Promise & The Peril

The Promise

LLMs are multilingual. An English-speaking analyst can now analyze Chinese, Japanese, or German filings instantly.

The Question

If you were analyzing a Chinese firm (e.g., Wuchangyu), which tool would you trust?

- **ChatGPT-4** (US-based)?
- **DeepSeek** (China-based)?

Poll: Which model likes Chinese public companies better?

The Reveal: "Foreign Bias"

Cao, Wang, Yi (2026)

We tested both models on ~5,000 Chinese firms.

Optimism

ChatGPT predicts higher prices, higher EPS, and generates rosier descriptions.

+12.5%

Higher Prices, EPS, Sentiment

Accuracy

Despite the optimism, ChatGPT is significantly LESS accurate on both prices and EPS.

+13.3%

Larger Price & EPS Errors

The Reversal

This reverses the classic human "Home Bias." Here, the foreigner is overly bullish.

Why? Information Asymmetry

The bias isn't from instruction tuning (e.g., "be nice to China")

The Hypothesis

ChatGPT is trained primarily on English/US internet data. US media coverage of Chinese firms is sparse and often misses local negative news (e.g., scandals, fines).

The Evidence

- **The Experiment:** Replace real Chinese firm names with fake names. Ask both LLMs to make predictions about these fictitious firms.
- **The Result:** No foreign bias!
- **The Implication:** Rules out instruction tuning / tokenization artifacts as the driver.

Case Study: 湖北武昌魚 (600275.SH)

CSRC Comment Letter

- Improper revenue recognition and related-party transactions
- Misleading disclosures in financials
- Mandate for corrective action and enhanced supervision
- Heightened stock Suspension risk

中国证券监督管理委员会 CHINA SECURITIES REGULATORY COMMISSION 湖北监管局

首页 | 证监局介绍 | 辖区监管动态 | 政务信息 | 办事服务 | 辖区数据 | 互动交流 | 专题专栏

当前位置：首页 > 政务信息 > 主动公开目录 > 证监局主题分类 > 行政执法

索 引 号	bm56000001/2021-00242175	分 类	行政执法行政监管措施
发布机构	湖北局	发文日期	2018年01月23日
名 称	湖北证监局关于对湖北武昌鱼股份有限公司采取出具警示函监管措施的决定		
文 号	[2018]2号	主 题 词	

湖北证监局关于对湖北武昌鱼股份有限公司采取出具警示函监管措施的决定

湖北武昌鱼股份有限公司：

经查，你公司在信息披露等方面存在以下问题：

根据你公司提供的十万头生猪养殖项目可行性研究报告，子公司鄂州市大鹏畜禽发展有限公司（以下简称大鹏畜禽）只是十万头生猪养殖项目（包括五个基地）中的一个基地，且以大鹏畜禽现实际拥有的猪舍2.30万平米及其他配套房屋0.25万平米推断，大鹏畜禽年产生猪最多不超过四万头。你公司在回复交易所多次问询时，均以大鹏畜禽年产十万头的结论回复上海证券交易所相关问讯并公告。经核查，大鹏畜禽年产十万头生猪缺乏依据，公司信息披露不准确。

上述问题违反了《上市公司信息披露管理办法》（中国证券监督管理委员会令第40号）第二条有关规定。根据《上市公司信息披露管理办法》五十九条有关规定，现对你公司采取出具警示函的措施。

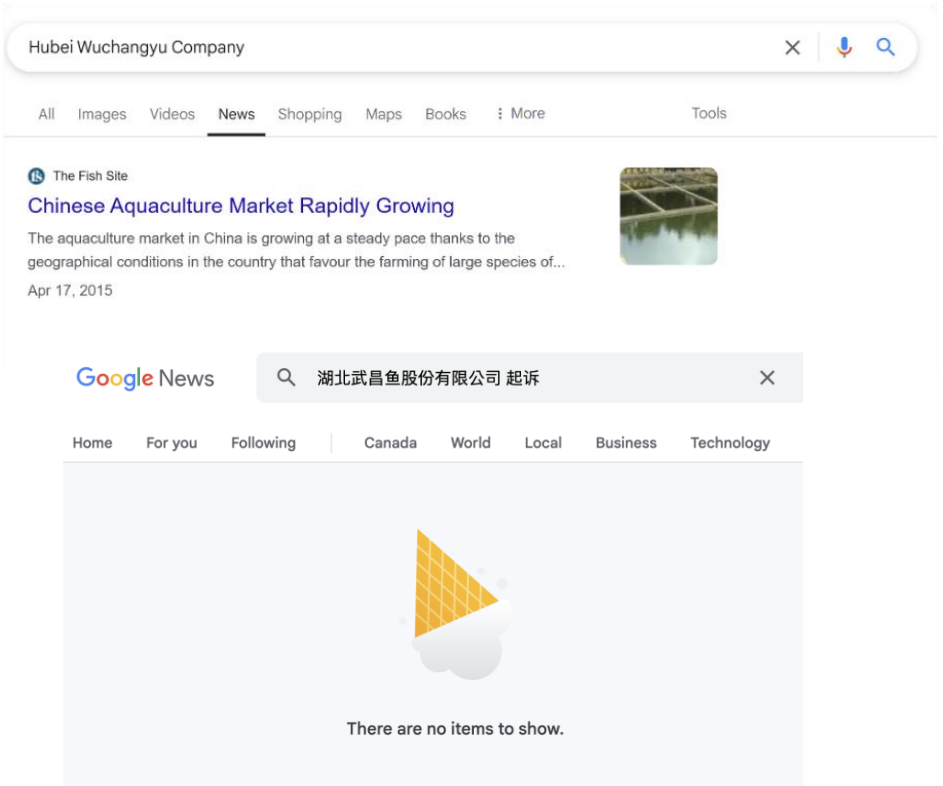
你公司应高度重视上述问题，及时采取有效措施，完善管理制度，优化信息披露流程。你公司应当自收到本决定书之日起30日内向我局提交书面报告。报告应当包括落实措施、预计完成时间、责任人等内容。

如果对本监督管理措施不服，可以在收到本决定书之日起60日内向中国证监会提出行政复议申请，也可以在收到本决定书之日起6个月内向有管辖权的人民法院提起诉讼。复议与诉讼期间，上述监督管理措施不停止执行。

Information Gap in Action

Chinese Media (百度)

US Media (Google News)



Why? Information Asymmetry

The bias isn't "political." It's informational.

The Hypothesis

ChatGPT is trained primarily on English/US internet data. US media coverage of Chinese firms is sparse and often misses local negative news (e.g., scandals, fines).

The Evidence

- **The News Gap:** US sources average 1.7 articles per Chinese firm vs. 26.9 from Chinese sources. The gap is worst for negative news.
- **Missing Bad News → Optimism:** ChatGPT's optimism most pronounced when US sources missing more negative coverage about the firm.
- **Missing Good News → Pessimism:** When US sources instead miss more positive coverage, ChatGPT becomes relatively more pessimistic.
- **News Injection Eliminates Bias:** Inject Chinese news into the prompt eliminates ChatGPT's relative optimism!

What Doesn't Fix It

Three alternative approaches.

✗

"Just let it Google things"

Web Search

Internet access does NOT eliminate bias. Signal-to-noise ratio low (relevant info drowned out) relative to news injection.

✗

"Newer models will fix it"

ChatGPT-5

Even a reasoning model is still biased. Not a reasoning problem, a corpus problem.

✗

"Prompt in their language"

Mandarin

Switching languages doesn't help. It's the corpus, not the prompt.

It's Not Just Price Targets

Ask the same model the same firm in three different ways — every answer leans the same direction.

YOU ASK

“What's your price target?”

STOCK PRICE

Higher targets for Chinese firms — even when the financial fundamentals are identical.

YOU ASK

“What will EPS be next period?”

EARNINGS FORECAST

Higher EPS forecasts for Chinese firms. Same direction, same population.

YOU ASK

“Describe the business.”

BUSINESS SENTIMENT

Rosier language about what the firm does. Not a forecast — a fundamental representation.

Not a forecasting glitch. The model's entire representation of these firms is rosier: across price forecast, earnings forecast, and the words it uses to describe the business.

Is the Bias Spreading to Humans?

1

US analysts get more optimistic about Chinese firms — Chinese analysts don't.

After ChatGPT's release, US analysts' EPS forecasts for Chinese firms became more optimistic relative to Chinese analysts. The shift is specific to Chinese firms — no parallel effect on US firms.

2

The shift tracks the news gap.

US analysts become more **optimistic** precisely for Chinese firms where US-language coverage is sparsest and missing more negative news. US analysts more **pessimistic** when missing more positive news!

3

Implication

As LLMs become (implicit) information intermediaries, their corpus biases could transfer to human decision-makers!

Key Takeaways

Yes … but the Mental Model is Incomplete

- **Yes:** if these are "thinking machines," the opportunity for white-collar work is enormous — speed, neutrality, and breadth applied to judgment-intensive tasks.
  - Good at extraction ("needle in the haystack") tasks, summarization, coding, multimodal

- **Yet:** the "mind" analogy is incomplete — an LLM is also a **statistical model**, trained then tuned. Each stage shapes capabilities *and* distortions:
  - **Stage 1 — Pretraining on a corpus:** develops representations of meaning and reasoning, but also encodes the patterns, associations, and biases of that corpus.
  - **Stage 2 — Instruction tuning (RLHF):** rewards helpfulness and task completion: reflect organizational preferences; pressure to "do too much."
  - **Stage 3 (optional) — Reasoning models:** trained on verifiable domains (math, code).
    Open question: do gains transfer to non-verifiable domains?

A Catalog of Documented Distortions

From the corpus (Stage 1):

- **Semantic**: multivalent terms ("core earnings") trigger nearby corpus associations and override the user's definition. *Shaffer and Wang (2026)*
- **Spatial:** geographic asymmetries in corpora. *Cao et al. (2026)*
- **Temporal:** future information leaks into predictions. *Levy (2026); He et al. (2025)*
- **Heuristic:** human heuristics (anchoring, representativeness) reproduced. *Bini et al. (2025)*

⇒ Common thread: distortions arise because the model learned its corpus.

From assistant tuning (Stage 2):

- **Reaching:** task-completion pressure produces output even when restraint is warranted.
- **Sycophancy & miscalibrated confidence:** outputs aligned to user expectations rather than truth.

Practical Takeaways for Harnessing LLMs

- **Don’t rely on the LLM for extensive arithmetic.** Use it to read text; call a tool for the math.
- **Avoid multivalent terms.** Choose clear, unambiguous terms: e.g., “recurring” not “core.”
- **If the reasoning path is clear, specify it.** Decompose into sequential prompts — don’t rely on a reasoning model to find the path itself.
- **Beware the unknown corpus.** If you’re not sure the requisite information is in the corpus, put it in the context window.
- **Beware the overly-helpful assistant.** Triangulate across models — different tuning regimes produce different blind spots.
- **Beware the confident and eloquent assistant.** Can’t use heuristic to approximate quality.

Questions?

The Automated Workflow

1. **Ingestion**
   NotebookLM
2. **Strategist**
   Qualitative Gem
3. **Assembler**
   Python/Excel Gem
4. **Analyst**
   Dupont/Apps Script
5. **Modeler**
   Valuation Gem

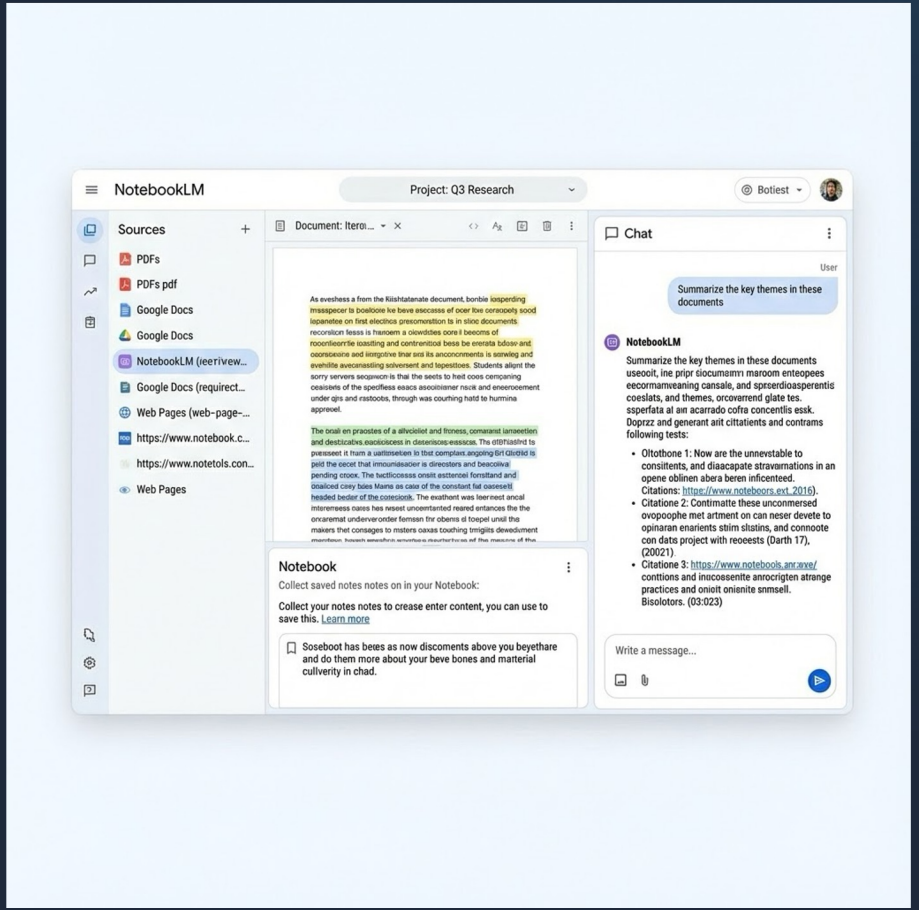
Step 1: Ingestion & Grounding

The Foundation of Truth

We begin by uploading the core financial documents into a dedicated NotebookLM notebook for the specific company.

- Latest 10-K (Annual Report)
- Recent 10-Qs (Quarterly Reports)
- Earnings Call Transcripts

This creates a closed knowledge base, ensuring all subsequent analysis is based strictly on company filings.



Step 2: "The Strategist" Gem

Role: Qualitative Bridge

This Gem acts as a Senior Strategy Consultant. It bridges the gap between qualitative strategy and quantitative inputs. It reads the MD&A and Call Transcripts to understand the *why* behind the numbers.

Key Outputs

- **Business Model Mechanics:** Revenue architecture and differentiation levers.
- **Success & Risk Factors:** Quantified metrics for tracking execution.
- **Scenario Framework:** Defines the specific assumptions for Bull, Base, and Bear cases.

Step 3: "The Assembler" Gem

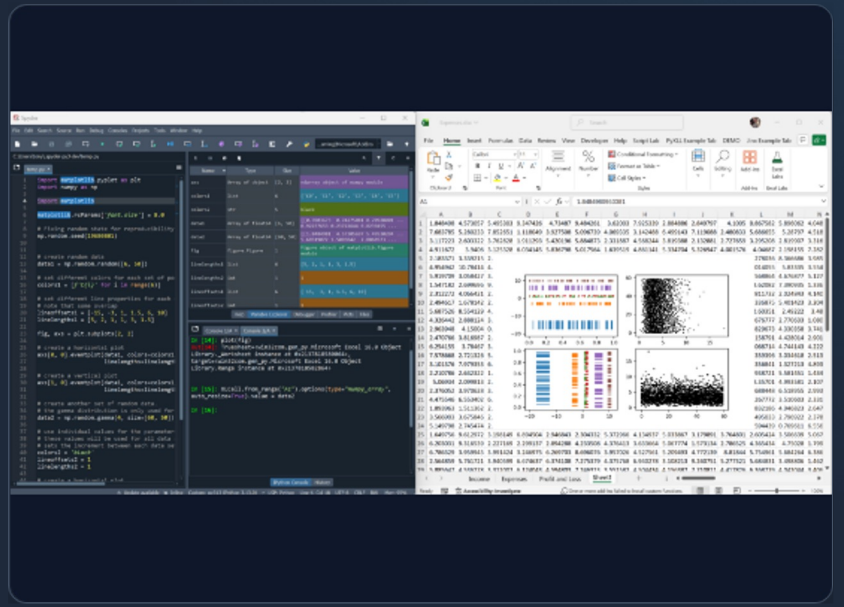
From PDF to Spreadsheet

This Gem is tasked with raw data extraction. It does not summarize. It generates a Python script to parse the 10-K and produce a multi-tab Google Sheet.

nb. No coding on your part! Just run it on Colab → Spreadsheets.

Core Functions:

- Extracts Income Statement, Balance Sheet, & Cash Flow Statement, Equity Statement, Core Earnings
- Normalizes sign conventions (e.g., expenses as negatives).
- Implements "Super-set Schema" logic to handle line item changes over years.



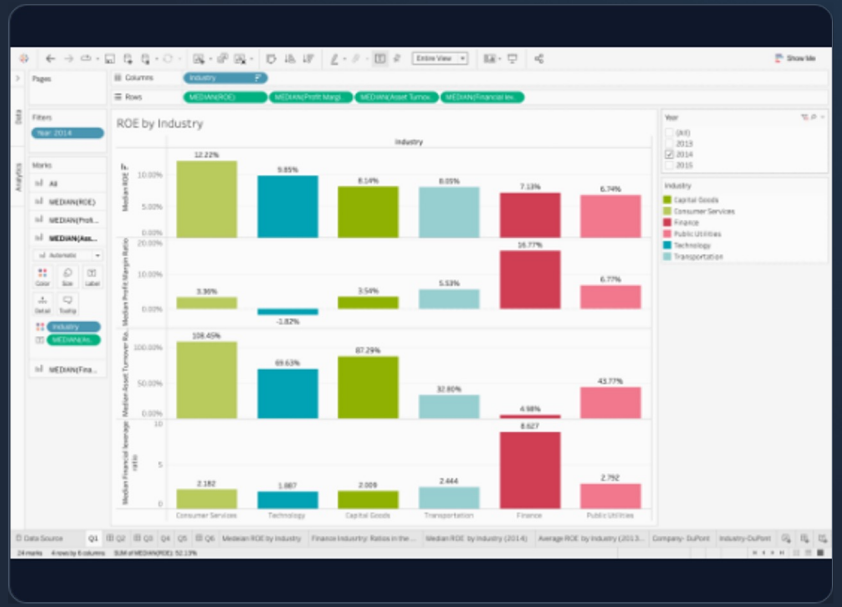
Step 4: "The Analyst" Gem

Automating Insight

Once the raw data is in Google Sheets, "The Analyst" Gem takes over. It writes Google Apps Script to perform rigorous financial calculations.

Key Deliverables:

- **Dupont Analysis:** Decomposes ROE into Net Profit Margin, Asset Turnover, and Financial Leverage.
- **Condensed Financials:** Aggregates complex accounting lines into standard operating models.
- **Audit Trail:** Flags any rows that don't balance.



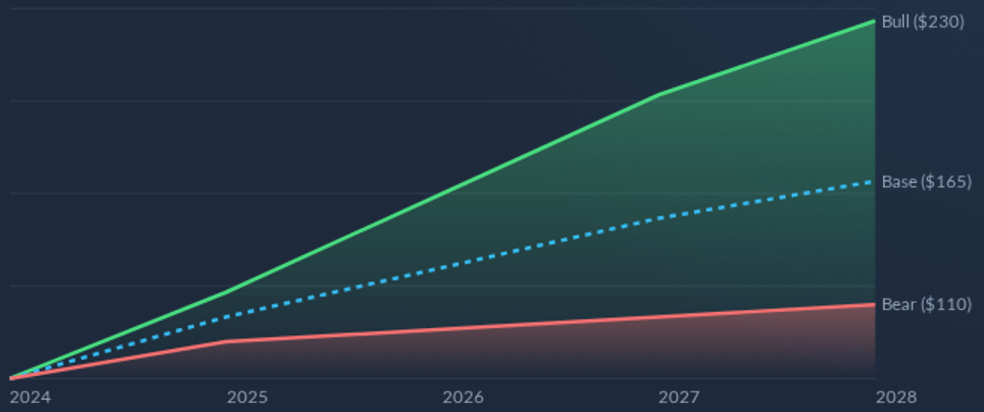
Step 5: "The Modeler" Gem

Valuation & Scenarios

The final Gem synthesizes the "Strategist's" framework with the "Assembler's" data.

It builds forecasts based on “bull”, “base”, and “bear” case scenarios.

Provides **DCF Valuation Models** with intrinsic value for each scenario.



The Final Deliverable

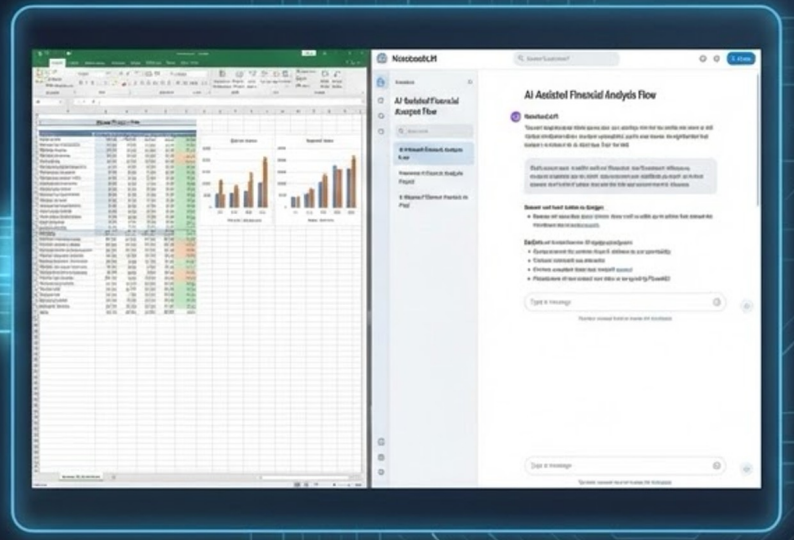
The Baseline Workbook + RAG

The output is not just a summary, but a functional tool.

Audit-Ready: Traces back to 10-K source data.

Live Models: Formulas are active, allowing the analyst to tweak growth assumptions instantly while querying Notebook LM.

Strategic Context: Includes the qualitative "Professor's Notes" generated by the Strategist Gem alongside the numbers.



What Can Go Wrong

Risk Awareness Checklist

Extraction Failures

Merged cells or poor PDFs can produce garbage. Mitigation: Always sanity-check extracted totals.

Hallucination Persists

Models may fill gaps with plausible nonsense. Mitigation: Verify material figures not directly cited.

Prompt Brittleness

Tuned on large-caps, may break on unusual filings. Mitigation: Expect iteration on new company types.

Code Bugs

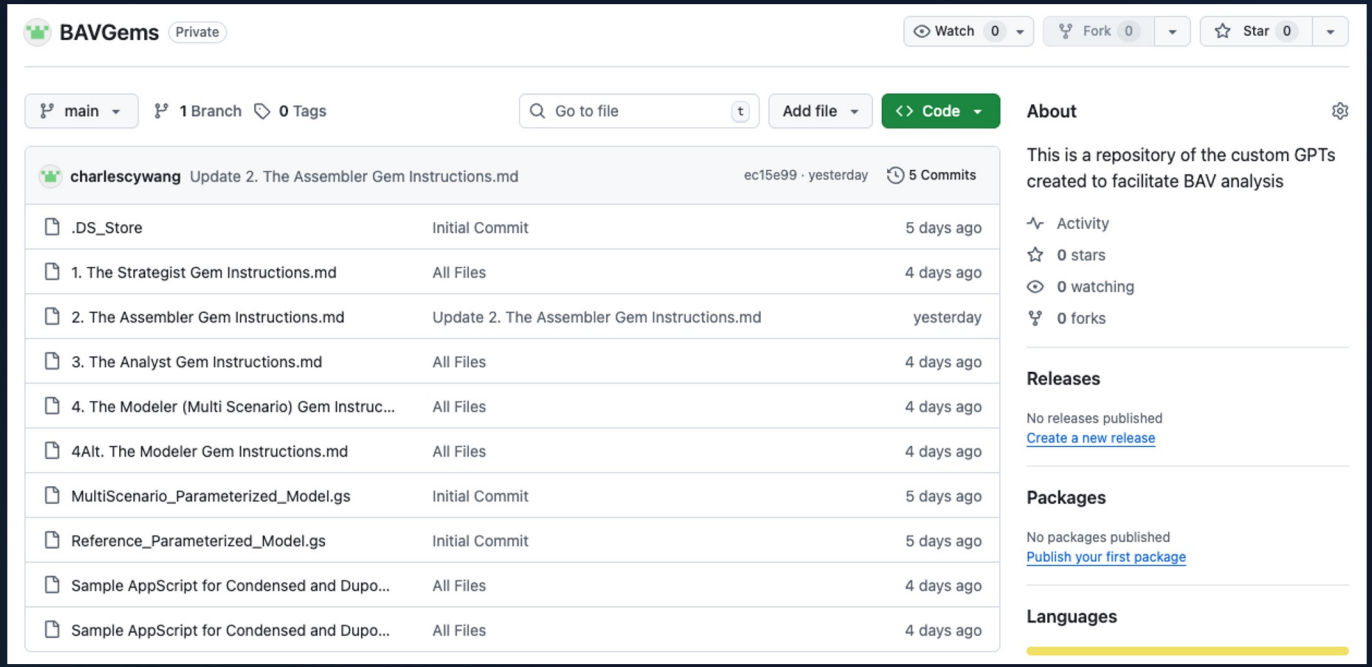
Scripts have bugs; sign errors propagate silently. Mitigation: Audit formulas & derived figures (e.g., ROE).

Automation Bias

Trusting polished output blindly is dangerous. Mitigation: You are the audit layer. Never skip review.

Resources: My “Gems”

System prompts available in repository:
https://github.com/charlescywang/BAVGems



Google Educational Resources

Intro to Gen AI

Basic Concepts & Applications

https://www.skills.google/course_templates/536

Intro to LLMs

Understanding Architecture

https://www.skills.google/course_templates/539

Prompt Design

Prompting Strategies

https://www.skills.google/paths/118/course_templates/976