



XBRL/CRAFT CONFERENCE  
PRESENTATION 2026

---

# CRAFT

Center for Research towards  
Advancing Financial Technologies

Advancing safer, smarter, stronger financial systems.

CRAFT@Stevens.edu 

201-216-3313



# What is CRAFT?

CRAFT is a multi-university consortium dedicated to shaping the public-good infrastructure of tomorrow's financial system. Through foundational research in AI governance, systemic resilience, digital market infrastructure, and secure data standards, we deliver the tools, insights, and talent that inform financial policy and make markets safer, smarter, and stronger.

➔ Federally Funded

➔ Industry/Academia Consortium

➔ Fintech Research

➔ First Center of it's Kind

# How It Works

CRAFT operates as a collaborative cycle where industry partners and academic researchers jointly identify, fund, and advance fintech research aligned with real-world needs. Projects are proposed, refined, and selected by members, then developed through ongoing collaboration into tangible outputs like white papers, intellectual property, and practical applications.



## Faculty Submit Fintech Research Proposals

Faculty submit concise pre-proposals aligned with industry needs, which are developed into structured research proposals leveraging academic expertise and collaboration.



## Proposals Are Voted On By Industry

Proposals are pitched to industry members, who evaluate, refine, and vote on projects to help shape research priorities.



## Winning Proposals are funded

Top-ranked proposals are approved and funded, with research teams collaborating with industry partners and providing ongoing updates and results.



## Industry can access IP, Student Talent

Members gain access to research outputs and engage with faculty and students, with opportunities to sponsor projects, recruit talent, and help bring innovations into real-world use.

# 3 Main Pillars

CRAFT is built on three core pillars that together create a unique value proposition for industry partners.

Our research advances the technologies shaping financial markets, our workforce initiatives develop the next generation of talent, and our thought leadership efforts influence the policies and frameworks that guide the industry forward.

By integrating these pillars, CRAFT provides members with a strategic platform to innovate, collaborate, and lead in an increasingly complex financial landscape.



## → Research

CRAFT drives collaborative, pre-competitive research across AI, financial infrastructure, digital assets, and systemic risk. Industry members work directly with leading faculty and researchers to develop technologies and frameworks that address real-world financial challenges and opportunities.

## → Workforce Development

CRAFT builds a direct pipeline to top-tier talent through student-led research, internships, and hands-on industry collaboration. Members gain early access to highly trained graduate students and future hires with expertise aligned to evolving financial technology needs.



## → Thought Leadership

CRAFT serves as a neutral platform for dialogue between industry, academia, and regulators. Through workshops, publications, and advisory engagement, members contribute to shaping policy, standards, and the future direction of financial systems.



# Performance

## Research & Innovation

- 24 Funded Research Projects
- Over \$2 Million Dollars for Research
- 125+ Research Proposals Submitted
- Phase 1 of NSF Program Successfully Completed
- Policy & Regulatory Engagement
- Focus Areas: AI, Language AI, Crypto Assets, DeFi, Quantum, Cybersecurity, Green Finance

## Industry & Academia

- 21 Industry Partners
- 3 Participating Universities
- 75 Academic Faculty
- 160+ Students
- 200+ Professionals Engaged Annually
- 5 Industry Inspired Workshop

# Performance

## Publications & Knowledge Output

→ 90 Individual Project Reports

→ 40 Project Presentations

→ 12 White Papers

→ 23 Conference Papers

→ 7 Journal Articles

## Talent & Workforce Development

→ Over 160 Students Supported

→ 20+ Students Active Annually

→ 35+ Talent Network Students

→ Direct Access to Student Talent

→ Students Placed at Top Firms

## Commercialization, Impact, & Growth

→ 3 Invention Disclosures

→ 1 Patent Pending

→ \$490K Annual Industry Funding

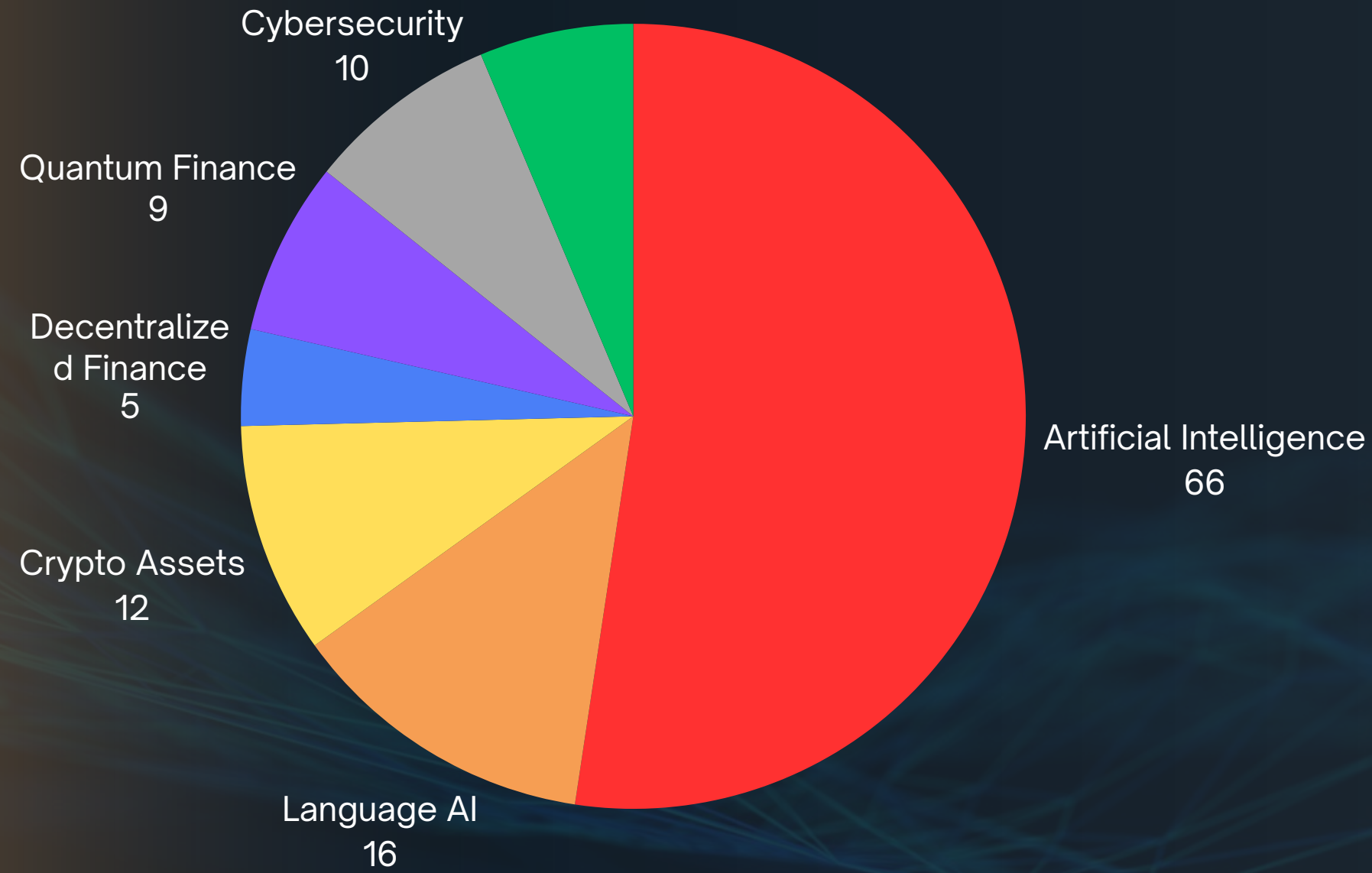
→ \$300K Annual NSF Funding

→ Partnerships with United Nations & NJFAST

# Research Proposals

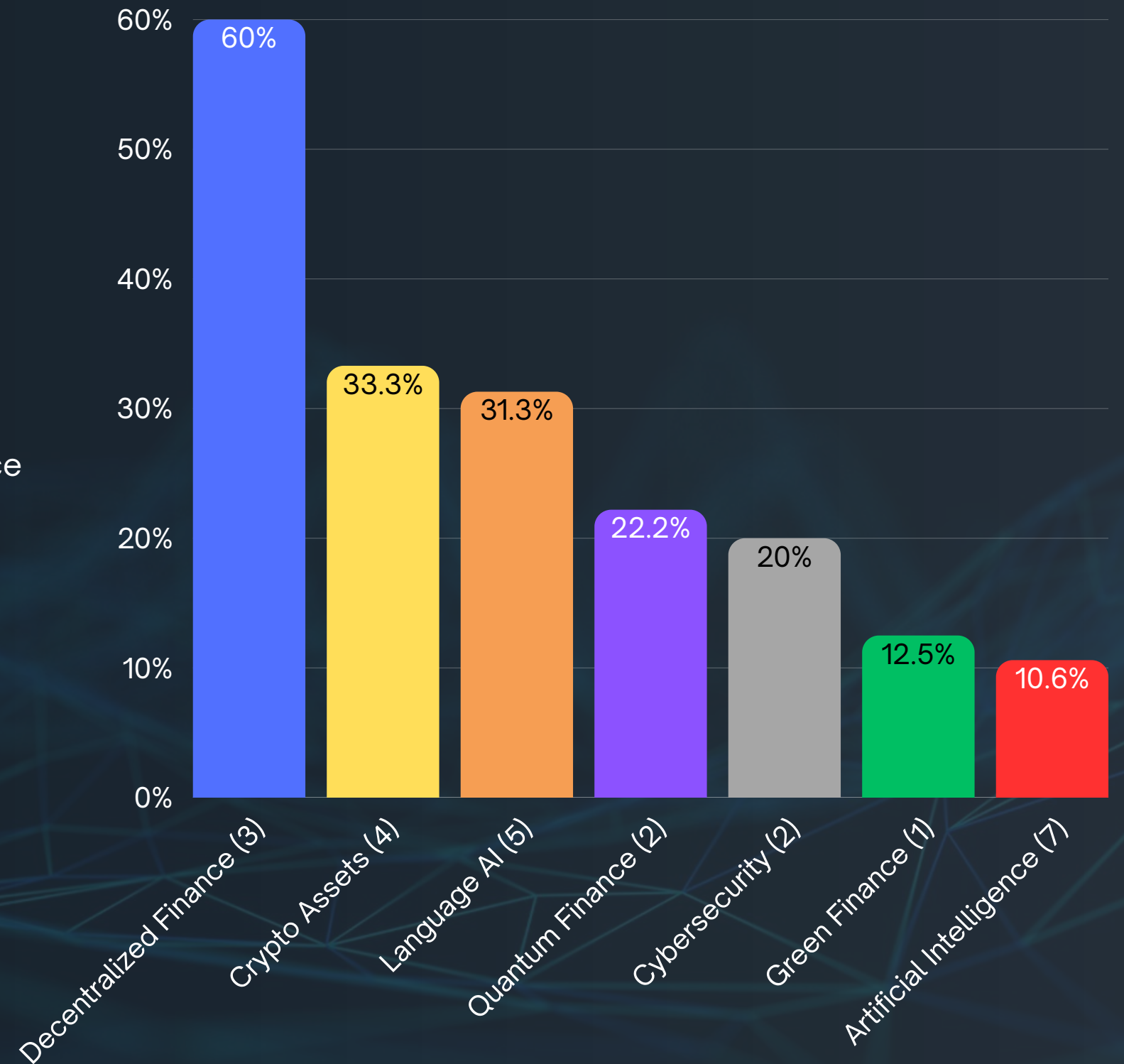
## Total Proposals by Theme

Total Proposals - 127



## Success Rate by Theme

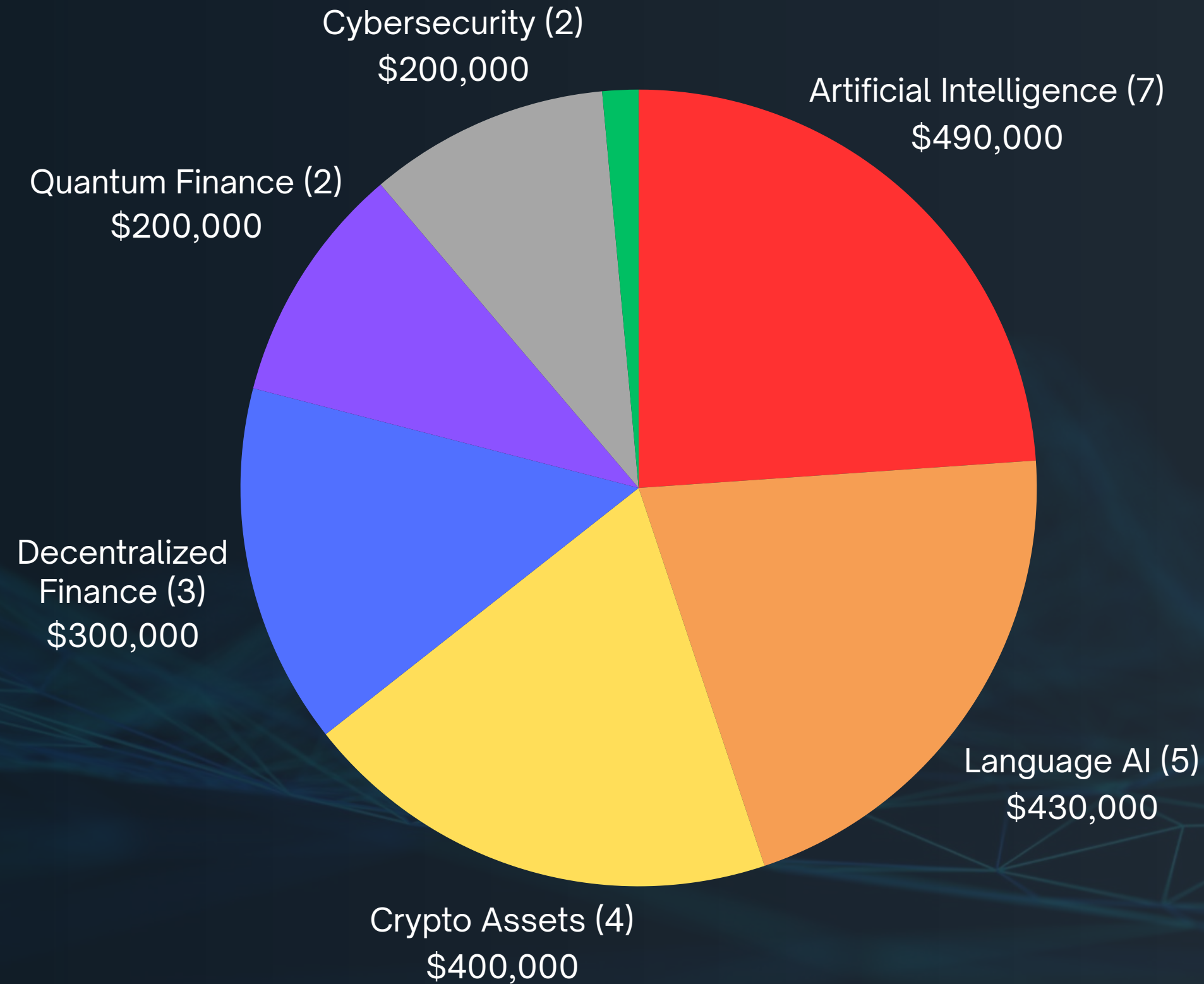
Percentage of Proposals Funded



# Research Portfolio by Theme

## Funded Projects by Theme

Total Funding - \$2,050,000



# Types of Membership



## Full Member

**\$50,000**

Geared toward large organizations  
Full vote on Industry Advisory Board  
Royalty-free access to center IP  
Leadership eligibility within IAB  
Direct influence over research priorities



## Associate Member

**\$25,000**

Geared toward mid-size organizations  
Half vote on Industry Advisory Board  
Royalty-free access to center IP  
Participation in strategic planning



## Affiliate Member

**\$5,000**

Geared toward start-up organizations  
Observer status in IAB meetings  
Access to workshops and research briefings  
Networking and collaboration opportunities



## Government/Lab/Network

**FREE**

Geared toward non-profits  
Custom engagement structures available

# Why Members join CRAFT

Membership in CRAFT provides organizations with direct access to innovation, talent, intellectual property, and strategic collaboration across the financial ecosystem. Through the NSF's IUCRC model, members leverage their investment to gain early access to cutting-edge research, participate in shaping the research agenda, and collaborate with leading academics, industry peers, and regulators.

In addition, CRAFT enables members to de-risk emerging technologies while building a strong pipeline of future talent. With access to royalty-free research outputs, hands-on workforce development opportunities, and a neutral platform for policy and regulatory engagement, members are positioned to stay ahead of industry change and help shape the future of financial systems.

→ Access to Talent

→ Access to IP & Results

→ De-Risk Innovation

→ Research Leverage

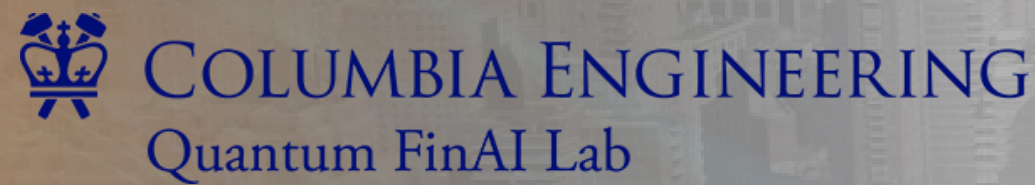
→ Policy & Regulatory  
Engagement

→ Workforce Development

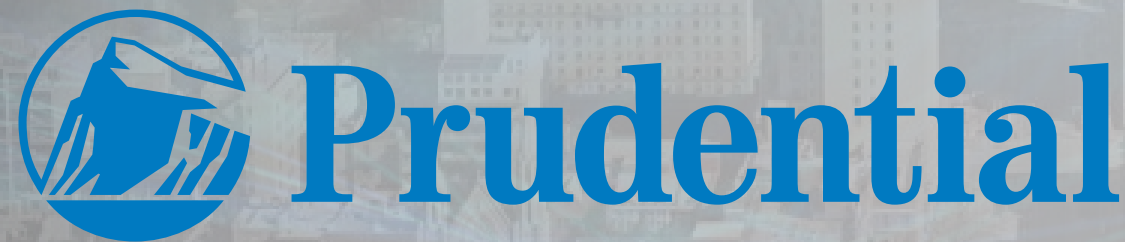
# CRAFT Industry Partners



BALCONY



Kennedys



WELLS FARGO





CENTER FOR RESEARCH TOWARD ADVANCING FINANCIAL TECHNOLOGIES

PRESENTATION 2026

---

# Thank You For Your Attention



201-216-3313



CRAFT@Stevens.edu





Julie Marlowe, CPA  
Office of Structured Disclosure (OSD)  
U.S. Securities and Exchange Commission

# INTERSECTION OF AI AND DATA

XBRL US & STEVENS INSTITUTE OF TECHNOLOGY  
MAY 15, 2026



# Disclaimer

*This presentation is provided in the author's official capacity as the Commission's Assistant Director of the Office of Structured Disclosure (OSD) but does not necessarily reflect the views of the Commission, the Commissioners, or other members of the staff.*



# Agenda

- What Does OSD Do?
- Data & AI
- Staff Observations and Guidance on Data Quality
- Free SEC Data Sets
- Statistics & Data Visualization
- Resources

# What Does OSD do?

- Make data accessible and easy to use
- Ensure data optimization and conformation to current market standards
- Design applications for conducting analyses
- Data collections & usage by publishing data sets and designing taxonomies and validation rules
- Market engagement

## Structured Data

The SEC is committed to making data more accessible and easier to use. The Division of Economic and Risk A other SEC divisions and offices to design data structuring approaches for required disclosures. DERA support data usage by designing taxonomies, validation rules, data quality assessments, and tools for conducting dat investors, regulated entities, and the public to support the submission and use of structured data.



[What Is Structured Data?](#)



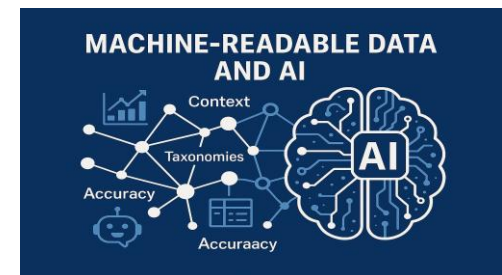
[History and Rulemaking](#)



[Data Quality Matters – Public Float Tagging Errors](#)

# Machine-Readable Data and AI

- AI driven by math and statistics
- Machine-readable data provides **CONTEXT**, making AI more accurate
- Data quality matters
- Filers define their data. Reduces interpretation by data users.
- AI chatbots using EDGAR API data
- Dept. of Commerce's [guidance](#): AI-ready data is enriched with **CONTEXTUAL** metadata



# AI Example: Research & Development Expense (per search on 5/12/26)

**Amazon spent \$108.521 billion on R&D in 2025, up from \$88.544 billion in 2024.**

## Detailed Breakdown

### 2025 R&D Expenses

- Amazon's total research and development expenses for the fiscal year ending December 31, 2025 were **\$108.521 billion**, representing a **22.56% increase** compared to 2024

# Amazon's 10-K, No R & D Expense

AMAZON.COM, INC.  
CONSOLIDATED STATEMENTS OF OPERATIONS  
(in millions, except per share data)

|  | Year Ended December 31, |            |            |
|--|-------------------------|------------|------------|
|  | 2023                    | 2024       | 2025       |
| Net product sales  | \$ 255,887              | \$ 272,311 | \$ 296,266 |
| Net service sales  | 318,898                 | 365,648    | 420,658    |
| Total net sales  | 574,785                 | 637,959    | 716,924    |
| Operating expenses:  |                         |            |            |
| Cost of sales  | 304,739                 | 326,288    | 356,414    |
| Fulfillment  | 90,619                  | 98,505     | 109,074    |
| Technology and infrastructure                                      | 85,622                  | 88,544     | 108,521    |
| Sales and marketing  | 44,370                  | 43,907     | 47,129     |
| General and administrative   | 11,816                  | 11,359     | 11,172     |
| Other operating expense (income), net                              | 767                     | 763        | 4,639      |
| Total operating expenses   | 537,933                 | 569,366    | 636,949    |
| Operating income   | 36,852                  | 68,593     | 79,975     |
| Interest income  | 2,949                   | 4,677      | 4,381      |
| Interest expense   | (3,182)                 | (2,406)    | (2,274)    |
| Other income (expense), net  | 938                     | (2,250)    | 15,229     |
| Total non-operating income   | 705                     | 21         | 17,336     |
| Income before income taxes   | 37,557                  | 68,614     | 97,311     |
| Provision for income taxes   | (7,120)                 | (9,265)    | (19,087)   |
| Equity-method investment activity, net of tax                      | (12)                    | (101)      | (554)      |
| Net income   | \$ 30,425               | \$ 59,248  | \$ 77,670  |
| Basic earnings per share   | \$ 2.95                 | \$ 5.66    | \$ 7.29    |
| Diluted earnings per share   | \$ 2.90                 | \$ 5.53    | \$ 7.17    |
| Weighted-average shares used in computation of earnings per share: |                         |            |            |
| Basic  | 10,304                  | 10,473     | 10,656     |
| Diluted  | 10,492                  | 10,721     | 10,827     |

# Machine-Readable Data & AI Use Cases



**Tagging vendor:**  
benchmark,  
draft  
narratives  
from tabular  
data,  
historical RF



**Audit firm:** Use  
machine-  
readable  
data to  
identify  
anomalies



**Data aggregator:**  
Integrated AI  
into reporting  
by using  
machine-  
readable data



**Per filer:**  
AI  
depends  
on solid,  
well-  
organized  
data from  
the start



**SEC: AI  
Task  
Force**

# SEC Use Cases

- Commission staff are building a semantic layer on top of the raw data assets in the Commission's enterprise data platform, transforming multiple complex data assets to enhance access and usability by staff using AI solutions<sup>1</sup>
- Comment Letter Review and Analysis<sup>2</sup>
  - Comment letter summaries
  - Enhances the review of public comment letters

1. "SEC Open Data Plan" available at [https://www.sec.gov/files/2025\\_sec\\_opendataplan.pdf](https://www.sec.gov/files/2025_sec_opendataplan.pdf)
2. SEC AI Use Case Inventory available at <https://www.sec.gov/ai>

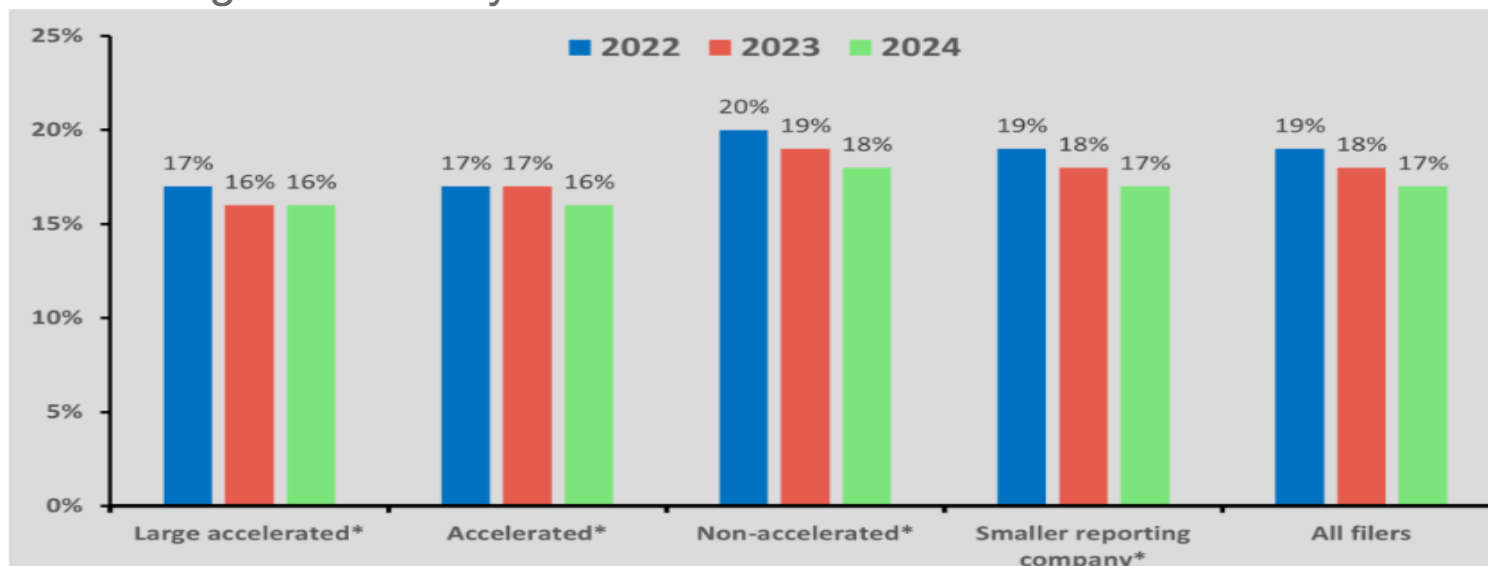


# Staff Observations and Guidance

- Data quality reminders: scaling errors, incorrect use of outdated tags, inappropriate use of custom tags, etc.

| Data Quality Reminders   | Last Update       |
|--|-------------------|
| <a href="#">Inappropriate Use of Consolidated Entities [Axis]</a>  | March 19, 2026    |
| <a href="#">Form 1-A Disclosure Inconsistencies</a>  | December 19, 2025 |
| <a href="#">Public Float Tagging Errors</a>  | July 21, 2025     |
| <a href="#">Incomplete and Inappropriate Tagging for Business Development Companies' Schedule of Investments</a> | June 16, 2025     |

- Custom tags trend analysis for GAAP filers



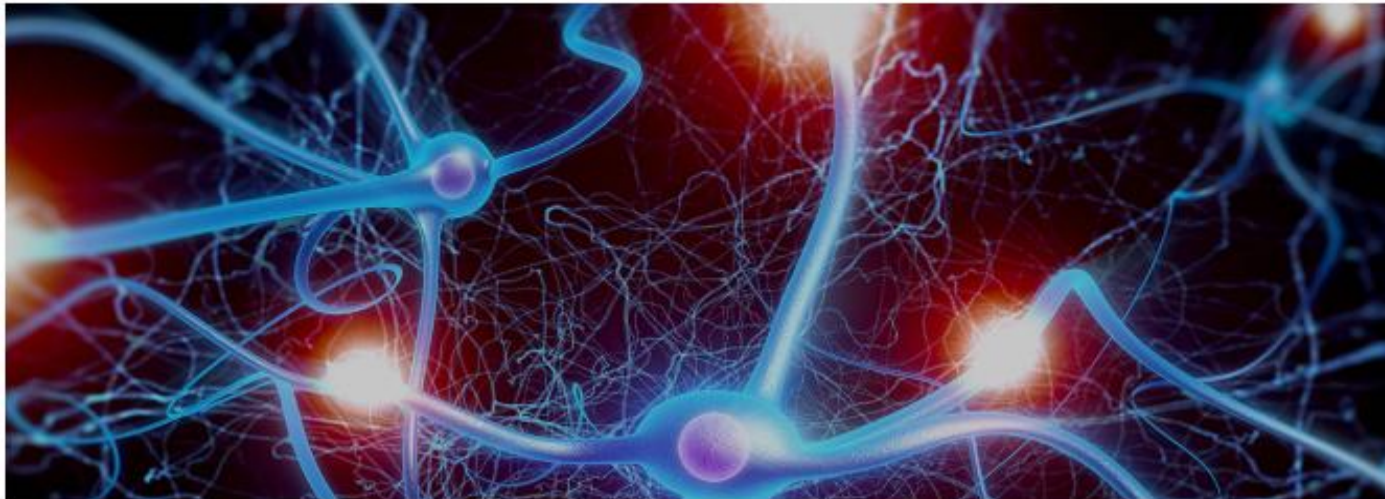
# SEC.gov | DERA Data Library

| <b>Data Set</b>  | <b>Last Updated</b> |
|--|---------------------|
| <a href="#">Business Development Company</a>               | May 6, 2026         |
| <a href="#">Crowdfunding Offerings</a>                     | April 6, 2026       |
| <a href="#">Financial Statement</a>                        | April 13, 2026      |
| <a href="#">Financial Statement and Notes</a>              | May 6, 2026         |
| <a href="#">Form 13F</a>                                   | March 2, 2026       |
| <a href="#">Form D</a>                                     | April 6, 2026       |
| <a href="#">Form N-CEN</a>                                 | April 6, 2026       |
| <a href="#">Form N-MFP</a>                                 | April 8, 2026       |
| <a href="#">Form N-PORT</a>                                | April 6, 2026       |
| <a href="#">Insider Transactions</a>                       | April 7, 2026       |
| <a href="#">Mutual Fund Prospectus Risk/Return Summary</a> | April 7, 2026       |
| <a href="#">Number of EDGAR Filings by Form Type</a>       | May 6, 2026         |
| <a href="#">Regulation A</a>                               | April 6, 2026       |
| <a href="#">Transfer Agent</a>                             | May 6, 2026         |
| <a href="#">Variable Insurance Product</a>                 | April 7, 2026       |

# Financial Statement and Notes Data Sets

<https://www.sec.gov/dera/data/financial-statement-and-notes-data-set.html>

January 2009 - April 2026



The Financial Statement and Notes Data Sets below provide the text and detailed numeric information from all financial statements and their notes. This data is extracted from exhibits to corporate financial reports filed with the Commission using eXtensible Business Reporting Language (XBRL). As compared to the more compact [Financial Statement Data Sets](#) which provide only the numeric information from face financials, the Financial Statement and Notes Data Sets provide significantly more disclosure data. The

# Python Code Examples: SEC F/S Data Sets

[sec-gov/python-for-dera-financial-datasets](https://github.com/sec-gov/python-for-dera-financial-datasets)

The screenshot shows the GitHub interface for the repository 'python-for-dera-financial-datasets' by 'sec-gov'. The repository is in the 'main' branch, has 2 branches, and 0 tags. The repository description is 'Python Code Examples for Accessing and Analyzing SEC's XBRL Data Sets'. The repository contains several files and folders, including 'examples', 'LICENSE', 'Populate\_Data.py', 'ReadMe.md', 'code-inventory.json', and 'requirements.txt'. The repository also has a README file and a CC0-1.0 license.

Code Issues Pull requests Actions Projects Security and quality Insights

main 2 Branches 0 Tags Go to file Code

sec-gov Merge pull request #2 from sec-gov/sec-patch-2 277a8a3 · 3 months ago 12 Commits

|                     |   |               |
|---------------------|---|---------------|
| examples            | Add files via upload                                | 3 months ago  |
| LICENSE             | Update license, add requirements.txt                | 10 months ago |
| Populate_Data.py    | Update examples from 2024_02_notes to 2024q1_notes. | 7 months ago  |
| ReadMe.md           | Update examples from 2024_02_notes to 2024q1_notes. | 7 months ago  |
| code-inventory.json | Add code inventory file.                            | 10 months ago |
| requirements.txt    | Update license, add requirements.txt                | 10 months ago |

README CC0-1.0 license

## Python Code Examples for Accessing and Analyzing SEC's XBRL Data Sets

This repository contains Python code examples to demonstrate how to read data from SEC's Financial Statement and Notes Data Sets using the Pandas package <https://pypi.org/project/pandas>. These examples help Python programmers understand different ways to access, analyze and visualize data from SEC's Data Sets.

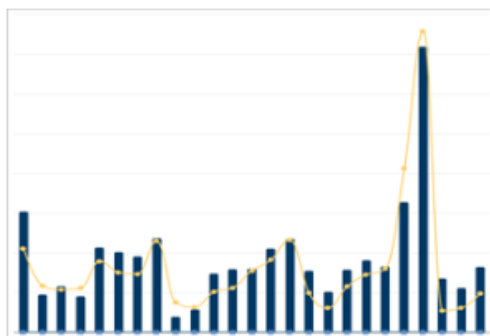
# Statistics & Data Visualizations

## [SEC.gov | Statistics & Data Visualizations](#)

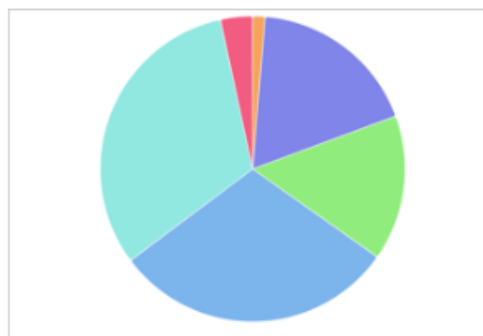
[Discover SEC Statistics and Data Visualizations](#) [🔗](#) (1-minute Introduction Video)

### Recent Data Visualizations

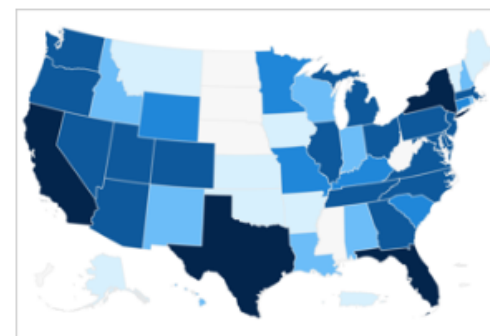
For all interactive data visualizations, visit [Data Visualization Gallery](#)



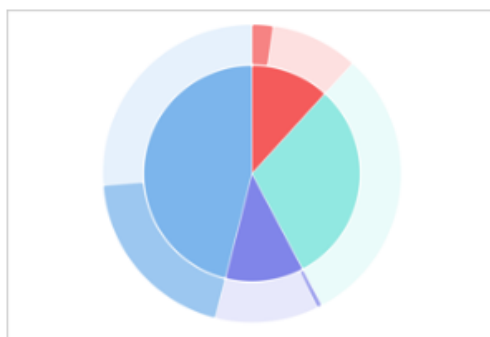
**Initial Public Offerings**  
**Number and Proceeds**  
(2000 - 2025 and 2000:Q1 - 2025:Q4)



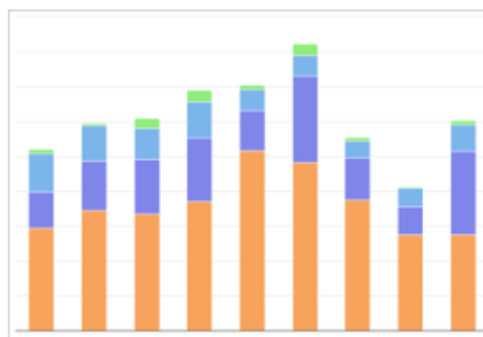
**Regulation D Offerings**  
**Total Amount Raised by Issuer Type**  
(2024 and 2025)



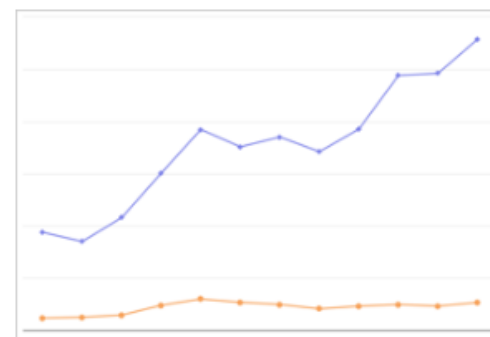
**Regulation CF Offerings**  
**Number of Offerings by Issuer Location**  
(May 16, 2016 - Dec. 31, 2025)



**Number of Reporting Issuers**  
**by Filer Status and Reporting Status**  
(2024)



**CMBS Issuances**  
**Number and Volume by Offering Type**  
(2016 - 2025 and 2016:Q1 - 2025:Q4)



**U.S. Households' Participation in Capital Markets**  
**Median and Mean**  
(1989 - 2022)

# Resources for More Information

Information on Structured Data:

[SEC.gov | Structured Data](https://www.sec.gov/structureddata)

Staff Observations and Guidance on Interactive Data Quality:

<https://www.sec.gov/structureddata/staffobsandguide>

Technical Questions on Structured Data:

[StructuredData@sec.gov](mailto:StructuredData@sec.gov)



Questions on Division of Corporation Finance's Interactive Data Rules and Compliance (select Office of Chief Counsel):

[SEC.gov | Corporation Finance Request Form for Interpretive Advice and Other Assistance](https://www.sec.gov/corpfinance/request-form)

Questions on Division of Investment Management's Interactive Data Rules and Compliance:

[IMOCC@sec.gov](mailto:IMOCC@sec.gov)

Announcements from the Office of Structured Disclosure: [SEC.gov | What's New](https://www.sec.gov/structureddata)

AI and Structured Data Forum:

# Optimizing Performance



May 15, 2026

## **Balancing the Global and the Local in the Development of a Machine-readable Data Infrastructure**

Alexi Aaltonen, Associate Professor  
*School of Business, Stevens Institute of Technology*

with Esko Penttinen, Aalto University, and Kalle Lyytinen, Case Western Reserve University

# Studying the Evolution of XBRL



eXtensible Business Reporting Language (XBRL) was introduced in 1999 and has evolved over decades into a global framework for creating machine readable financial statements data<sup>1</sup>

An early success case in the private sector is the use of XBRL at scale: today, regulatory mandating the use of XBRL instance documents of companies.

One of the authors (Esko Penttinen) has been active in local XBRL jurisdiction.

Ongoing historical study based on interviews, documents, participation in events, thesis work, etc. to identify and analyze critical junctures in the evolution of XBRL.

**Why did it take decades to make financial reporting data machine-readable? What can we learn from this?**

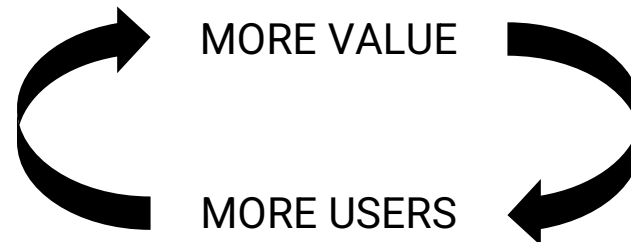
<sup>1</sup> We mostly focus XBRL Financial Reporting (not XBRL Global Ledger).

# What Makes Machine-readable Data Possible?



Large-scale **data infrastructures** allow data to be produced and shared between organizations in a machine-readable format.

Data infrastructures are sociotechnical arrangements that align standards, organizational IT capabilities, information producers and users across system and organizational boundaries so that data retain the same semantics across settings (Bygstad 2010, Hanseth and Lyytinen 2010).



*No one will share their data if there are no users, and users will not join the infrastructure if no one shares their data.*



# Research Question

Unlike enterprise systems, digital platforms, and applications, data infrastructures cannot be designed in a traditional sense because the solution space is not controlled by any single party.

To make a data infrastructure happen, one must solve two challenges (Hanseth & Lyytinen, 2010):

1. **Bootstrap challenge.** How to initially get necessary actors on board to make the infrastructure happen?
2. **Adaptability challenge.** How to adapt the infrastructure as new parties join and environment changes?

*How does a machine-readable data infrastructure solve the bootstrap and adaptability challenges?*

# What Does the Literature Say?



Information infrastructures literature suggests two strategies for resolving the challenges.

1. **Gateway-based architecture.** Technical translators facilitate communication between local implementations.
2. **Decentralized governance.** Infrastructure is cultivated bottom-up in local settings that manage their own implementations.

Yet, neither of these strategies is suitable for machine-readable data infrastructures.

1. Machine-readability is about (re)usability of data across boundaries without having to reconcile what the data means.
2. Machine-readability requires common semantics and formats, which are difficult to achieve through decentralized governance.

| INITIAL JUNCTURE   | FOLLOWING JUNCTURES  | DESCRIPTION   | CHALLENGES   |
|--|--|---|--|
| <b>1999</b><br><i>Decision to build XBRL on XML</i>  | 2020: The creation of the Open Information Model (OIM)   | INITIAL: XBRL was developed on XML and marked a strategic and future-proof decision as XML was already widely used and universally accepted across technical platforms allowing interorganizational linkages.<br>OVER TIME: The Open Information Model provides a syntax-independent model for XBRL data, allowing reliable transformation of XBRL data into other representations such as JSON and CSV.  | INITIAL: Addresses a bootstrap problem in technological architecture<br>OVER TIME: Addresses an adaptability problem in technological architecture     |
| <b>1999</b><br><i>Focus on statutory financial reporting</i>   | 2005: Extension to transactional data (XBRL GL)<br>2013: Extension to sustainability reporting   | INITIAL: Through its focus on statutory financial reporting, XBRL was able to attract the key stakeholders in its community to work on a digital reporting stream from companies to regulators.<br>OVER TIME: Extension to transactional data in form of XBRL GL allowed the community to involve software vendors in its development. The integrated reporting approach refers to the bundling of different financial and non-financial (e.g., sustainability reports) company reports under XBRL.   | INITIAL: Addresses a bootstrap problem in infrastructure governance<br>OVER TIME: Addresses an adaptability problem in technological architecture      |
| <b>1999</b><br><i>Focus on regulators and closed systems</i>   | 2014: Supporting open data initiatives   | INITIAL: Focusing on regulators in its early phases allowed for better control over development processes.<br>OVER TIME: To provide legitimacy on XBRL beyond a mere compliance act, International and some jurisdictions began to frame XBRL more explicitly in terms of open and reusable structured data.  | INITIAL: Addresses a bootstrap problem in infrastructure governance<br>OVER TIME: Addresses an adaptability problem in infrastructure governance       |
| <b>2002</b><br><i>The creation of a jurisdictional structure</i>   | 2006: The establishment of the XBRL Standards Board (XSB)<br>2009: The establishment of the Best Practices Boards (BPB)  | INITIAL: The governance structure in XBRL is organized around a network of independent, non-profit organizations known as jurisdictions, each jurisdiction being a member in XBRL International.<br>OVER TIME: The XSB is responsible for managing the production of the consortium's technical materials. The high-quality documentation ensures single point of contact for queries on, e.g., base specification and maintenance. The BPB is responsible for producing and disseminating information regarding successful implementations of XBRL specifications. The implementation guidance material is useful for implementers to avoid "reinventing the wheel" and stumbling on the same roadblocks as previous implementers.   | INITIAL: Addresses a bootstrap problem in infrastructure governance<br>OVER TIME: Addresses the tension on centralization vs. decentralization tension |
| <b>2003</b><br><i>Keeping the foundational specification 2.1 intact</i>  | 2006: The introduction of the dimensions specification<br>2009: Introduction of the formula specification<br>2010: The introduction of Inline XBRL (iXBRL)<br>2014: The creation of the table linkbase | INITIAL: The integrity of the foundational 2.1 specification has been maintained intact since its launch in 2003. The specification defines the basic building blocks of facts, instance documents, concepts, and taxonomies.<br>OVER TIME: Additional tools and linkbases introduced to respond to the requirements emerging from the community to (i) provide a standard mechanism for defining rules in a taxonomy that can be applied against instance documents (the formula specification), (ii) combine human-readable and machine-readable financial reporting into the same document (iXBRL), (iii) enable the reporting of multi-dimensional facts against dimensions defined in an XBRL taxonomy (Dimensions specification), (iv) provide a mechanism for taxonomy authors to define a tabular layout of facts to be used for both presentation and data entry (Table linkbase). | INITIAL: Addresses a bootstrap problem technological architecture<br>OVER TIME: Addresses an adaptability problem in technological architecture        |
| <b>2009</b><br><i>Decision taken by U.S. SEC (the U.S. Securities and Exchange Commission) to mandate XBRL for publicly listed companies</i> | 2008: Japan FSA<br>2011: Korea FSS<br>2011: HMRC Corporation Tax and Companies House<br>2014: COREP/FINREP reporting<br>2016: EIOPA Solvency II<br>2020: ESEF reporting                                | INITIAL: The U.S. SEC mandate in 2009 marked an important milestone being the first high-profile, large-scale filing program that mandated all publicly listed US-based companies to adopt XBRL for their yearly and quarterly statutory financial statements. The mandate had far-reaching effects on the adoption of XBRL worldwide.<br>OVER TIME: Spillover effect of the U.S. SEC mandate on other mandatory filing programs across the globe, both for statutory financial reports, but also for tax reporting and other government reporting.   | INITIAL: Addresses a bootstrap problem in infrastructure governance<br>OVER TIME: Addresses a bootstrap problem in infrastructure governance           |

# GOVERNANCE

## Hierarchical governance

Document receivers  
(regulators)

Companies

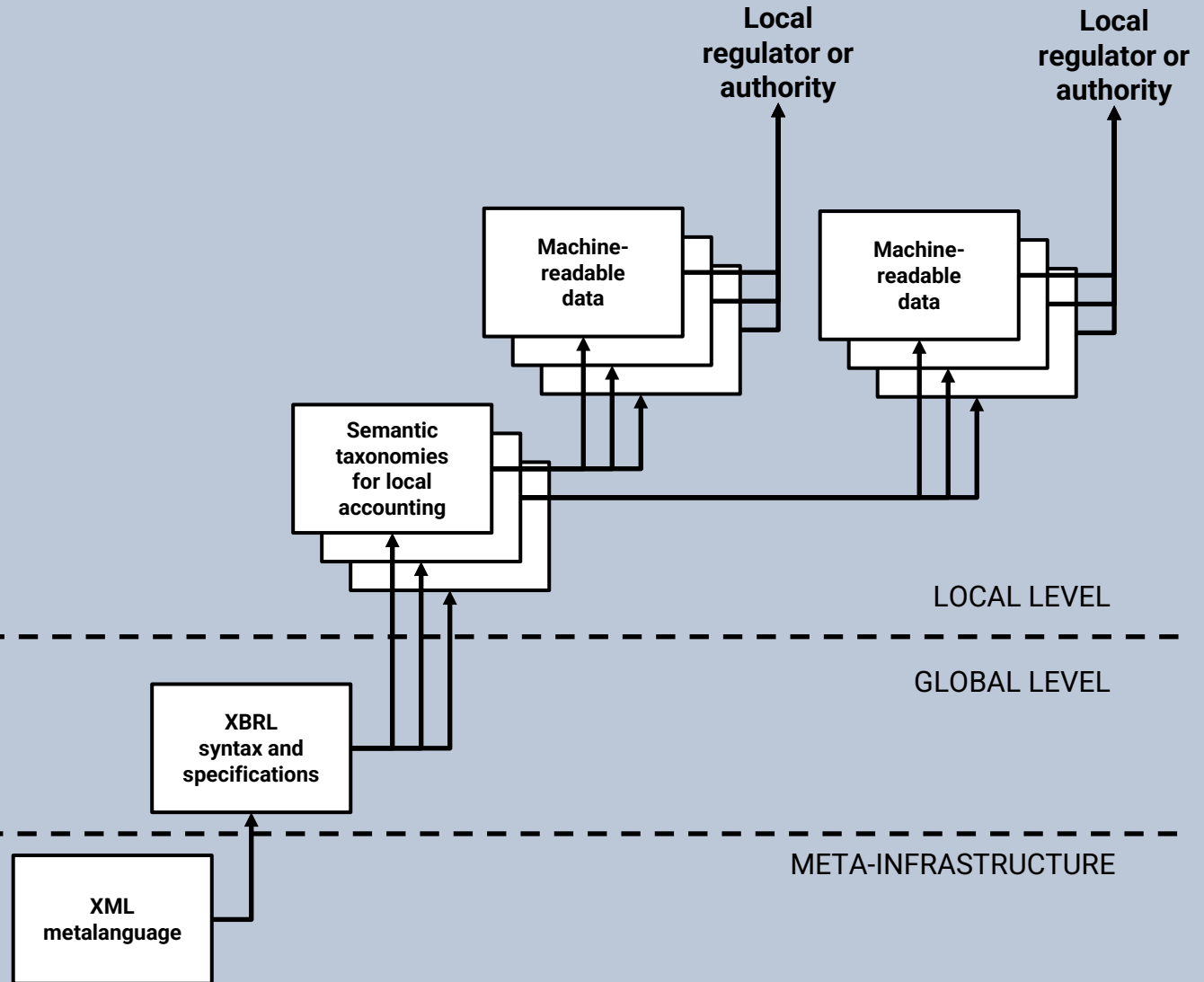
## Federated governance

XBRL jurisdictions  
and/or local  
authorities

XBRL  
International

W3C

# ARCHITECTURE





# Some Lessons Learned



It's not just about the model of governance (we see different models of governance in the case) but **how different levels of governance are aligned with the technical architecture.**

Getting the global—local divide right is critical both in governance and technical architecture. **Do not necessarily try to make the data themselves global** like timestamps or geolocation data. XBRL data are 'globally local'.

**The semantics of data need to be tied to adequate human institutions** that may not exist at global level (there are no global accounting semantics), yet the global syntax and meta-infrastructure are still critical.



# Structured Data in the Age of AI



Foundation models are trained on global data. Therefore, it is unclear what accounting standards they have ingested and ‘follow’.

Unstructured data, e.g., natural language is ambiguous—this is not a matter of more or less exact information extraction, but an inherent feature that makes human languages powerful.

It is not clear how one could ensure that relevant information has been disclosed unless it is reported in a structured format.

However, we probably want to make LLMs to understand accounting better. E.g., relevant standards may need to be fed as context for inference? This might be a new role for local jurisdictions.

Does the current alignment of XBRL governance and architecture still work with AI-driven environment and technology? What happens if the original XML meta-infrastructure is abandoned with open information model?

# Thank You!



<https://www.linkedin.com/in/aleksiaaltonen>

# Data

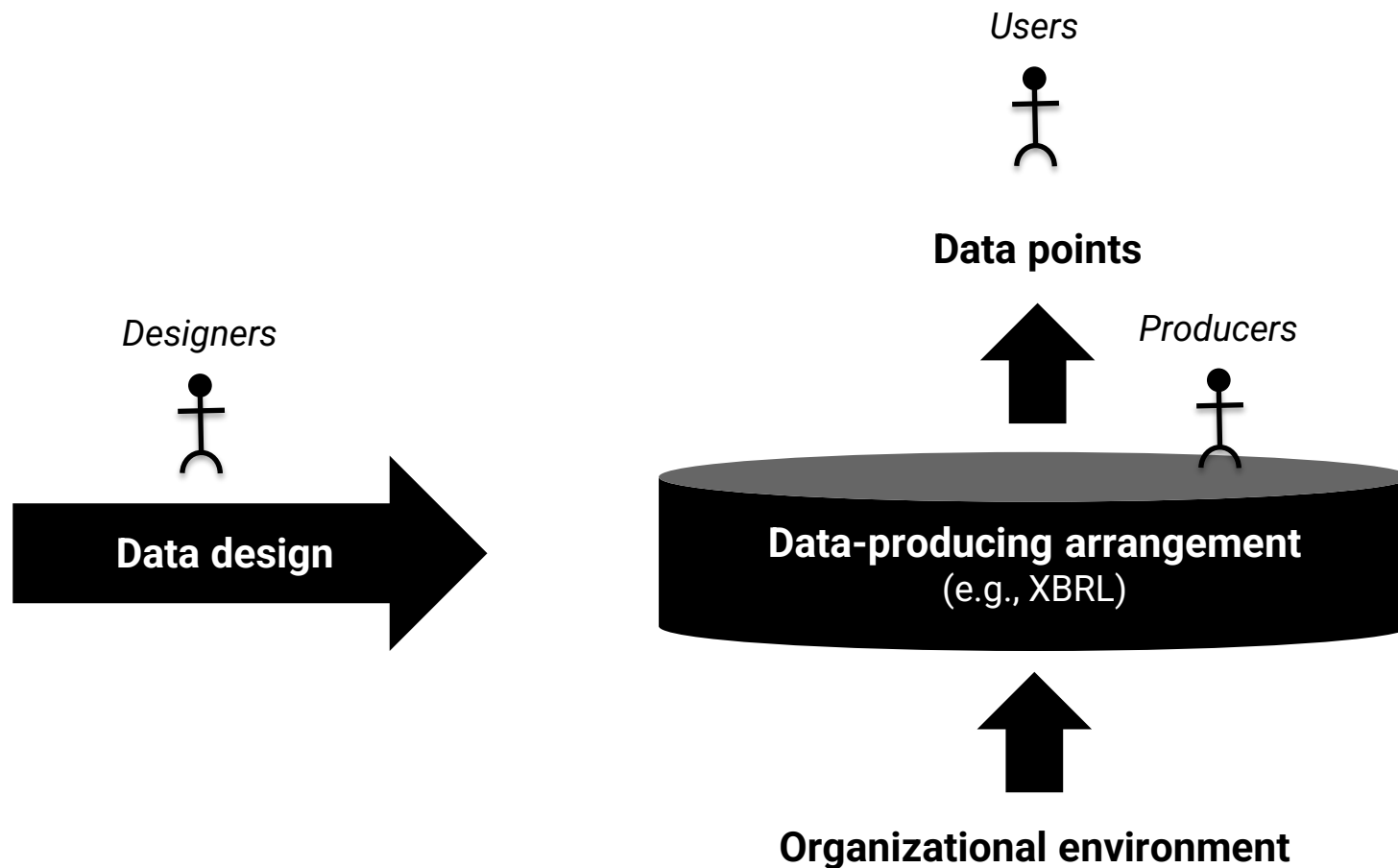


Data are the key resource  
in the AI-driven economy.

However,  
data are not oil.

Data are human-made material that  
are designed and produced under  
organizational circumstances.

# Data Design and Production



## **Can AI Reasoning Improve Accounting Estimation? Evidence from CECL Allowance**

Arion Cheong  
Stevens Institute of Technology

Sponsored by:

CRAFT (Center for Research toward Advancing Financial Technologies) / Park Avenue Finance

# Background: Current Expected Credit Loss (CECL)



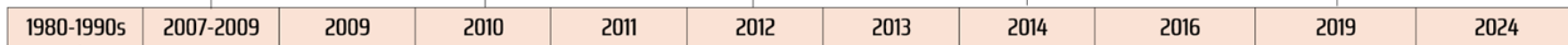
**Financial Crisis**  
Banks delayed loan loss recognition loan losses until they became uncollectible, contributing to procyclical bank lending during the 2007 financial crisis by postponing loss recognition during economic expansions and carrying losses forward to recessions.

**Initial FASB Exposure Draft**  
Recommended removal of the preexisting "probable" threshold for loss recognition.  
Recognition at each reporting date of an allowance for all full contractual-life expected credit losses.

**Final FASB Exposure Draft**  
Recommends the CECL model. Required recognition, at day one, of all expected losses over the full contractual life.  
Final FASB exposure draft, leading to Accounting Standards Update 2016-13.

**IFRS 9**  
Mainly included the elements of final IASB exposure draft.  
It addresses the accounting for financial instruments containing classification and measurement, impairment of financial assets and hedge accounting.

**ASU 2019-10**  
Extend the effective dates to be staggered between larger public companies and all other entities.  
Extended the adoption deadline by an additional year for private firms and SRC.



**Loss Overprovisioning**  
Different banks interpreted the same loan loss accounting standard differently, with most banks at the time interpreting the rule more in line with an expected loss model.  
SEC and FASB released guidelines re-emphasizing the "incurred loss" aspect.

**Initial IAS Exposure Draft**  
Recommended the recognition credit losses should use a less restricted forward-looking information set than that used by the pre-existing incurred loss-model.

**FASB/IASB JOINT DOCUMENT**  
Sought a compromise converged solution between FASB and IASB.  
Recognition: 12-month expected credit losses at initial recognition; lifetime ECL after significant credit risk increase.

**FINAL IASB EXPOSURE DRAFT**  
Recommended a three-bucket method, termed "Stage 1", "Stage 2" and "Stage 3".  
Final IASB exposure draft, leading to elements of IFRS 9.

**ASU 2016-13**  
Introduced the CECL Model. Effective for FY beginning after  
- Dec 15, 2019 (public business entities that are SEC filers).  
- Dec 15, 2020 (for all other SEC filers).  
- Dec 15, 2021 (for other entities).

**FASB's Proposed Update**  
Address the challenges when applying Topic 326 by private firms and NFP entities.  
Assume that current conditions as of BS date persist.  
Consider the collection activity after BS date for ECL estimation.

# Incurred Loss Model vs. CECL Model



| Aspect                            | Incurred Loss Model  | CECL Model  |
|-----------------------------------|--|---|
| <b>Governing Standard</b>         | FASB's ASC 450-20 and ASC 310-10   | FASB's ASC 326  |
| <b>Recognition Timing</b>         | Losses are incurred when “probable” and reasonably estimable for groups with similar risks. Losses are recognized only when there is evidence of impairment for large heterogeneous loans. | “At day 1” recognition (loan origination) (forward-looking).  |
| <b>Data Used</b>                  | Historical data and current observable conditions.   | Incorporate <b>reasonable and supportable forecasts</b> of future economic conditions.                |
| <b>Forecasting Method</b>         | Simple historical loss rate analysis.  | Complex, includes multiple economic scenarios.  |
| <b>Measurement Period</b>         | Short-term, incurred losses only.  | Long-term, Entire life-of-loan (full contractual term).   |
| <b>Time Horizon</b>               | Short-term (12-month loss emergence).  | Focuses on lifetime expected losses over the loan’s term.   |
| <b>Financial Impact</b>           | Lower reserves in good financial conditions. During economic downturns, provisions spike sharply.  | Higher upfront reserves reduce reported earnings at origination, but smoother provisioning over time. |
| <b>Role of Quant Analyst Team</b> | Validate historical loss models; backward-looking calibration.   | Develop forward-looking models, integrate macro forecasts, overlays, and scenario analysis.           |

# Expected Credit Loss Estimation



## 1. Industry Approach

- a. Loss rate methods: Snapshot, Weighted Average Remaining Maturity, Vintage methods.
- b. Non-loss rate methods: Roll-rate methods, probability of default, discounted cashflows.

## 2. Academic Approach

- a. Forward-looking measures outperform historical net charge-offs in one-year prediction with a cross-sectional regression (Harris et al., 2018).
- b. Bank stock prices capture lifetime losses beyond reported allowances under vintage analysis (Wheeler, 2021).
- c. Macroeconomic forecasts improve long-term loss predictions with a model that combines cross-sectional predictions and a dynamic factor model, which incorporates macroeconomic variables (Lu & Nikolaev, 2022).

# Critical Audit Matters (CAMs)



| Bank                | Annual Report Year | Excerpt   |
|---------------------|--------------------|---|
| Bank of America     | 2024               | “Also included in the allowance for loan and lease losses are qualitative reserves to cover losses that are expected but, in the Corporation’s assessment, may not be adequately reflected in the quantitative methods or the economic assumptions”   |
| JPMorgan Chase & Co | 2023               | “..allowance for loan losses...is a critical audit matter [given] the significant judgment and estimation by management in determining the quantitative calculation utilized in their credit loss estimates and the adjustments to take into consideration model imprecision, emerging risk assessments, trends and other subjective factors that are not yet otherwise reflected in the credit loss estimate |
| TD Bank             | 2023               | “Auditing the allowance for credit losses was complex..because of., the forward-looking nature of the key assumptions, ..in measuring the ECL. Key areas of judgment include evaluating:..the assumptions used in the ECL scenarios including forward-looking information, assigning probability weighting; and the assessment of the qualitative component applied to the modeled ECL                        |

# Research Gap: Qualitative Adjustments & Bounded Rationality



- Current estimation models cannot reflect qualitative adjustments.
  - Capture structured numeric relations only (Harris, Khan, & Nissim, 2018; Wheeler, 2021; Lu & Nikolaev, 2022).
  - Not capture qualitative reasoning that subject matter experts apply to the same data (e.g., non-linearities, bank-specific risk).
- Human experts face bounded rationality constraints (Simon, 1955, 192; Hogarth 1993; Libby et al. 2002).
  - Restricted information access (inability to process all available data),
  - Limited computational capacity (reliance on heuristics that introduce biases such as anchoring; Tversky & Kahneman, 1974), and
  - Difficulty evaluating hierarchical complexity.

# Qualitative Adjustments



## Definition

Management's adjustments to historical loss experience to reflect current conditions and reasonable and supportable forecasts not already captured in the quantitative estimation model **(ASU 2016-13)**.

## Challenges

Subjective in nature **(ASU 2016-13)**

No standard approach to quantify the qualitative adjustments **(OCC, 2022)**

No predefined set that should be considered as qualitative factors **(FDIC 2020)**

Costly and requires continuous monitoring **(FASB, 2025)**.



# Research Questions

- Background
  - As evidenced, regulators and auditors have expressed concerns regarding the reliability, consistency, and potential bias in these estimates.
  - These observations reflect increased managerial discretion and estimation uncertainty
- Research Question: Can AI reasoning improve the estimation of allowance for credit losses (ACL) under the current expected credit losses (CECL) model by augmenting traditional methods?

# Research Goals

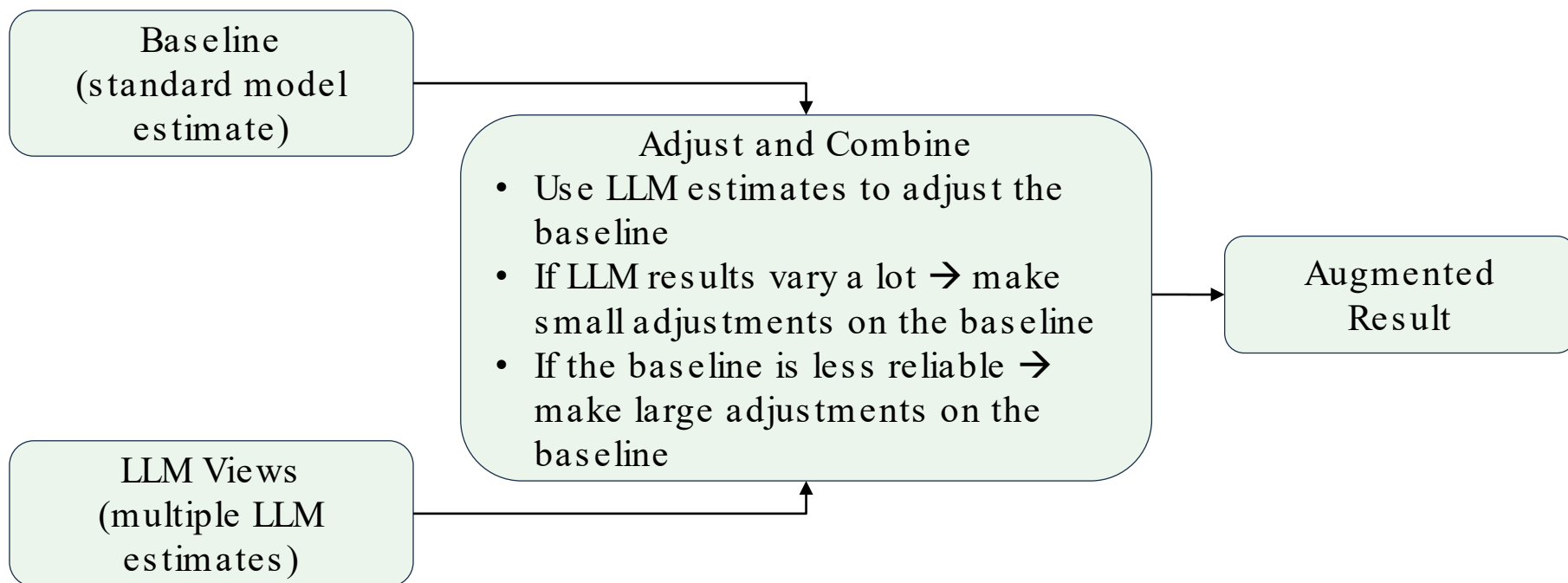


- Goals:
  - Task 1: Develop a Bayesian shrinkage framework that combines econometric baselines with LLM expert views, providing a theoretically grounded approach to qualitative adjustments.
  - Task 2: Fine-tune LLMs via Group Relative Policy Optimization (GRPO) to improve the quality of LLM expert views within the framework.
  - Task 3: Evaluate robustness across different pre-trained LLMs and test for look-ahead bias.

# Bayesian Shrinkage: Augmenting the Baseline



**Motivation:** CECL requires forward-looking adjustments to baseline credit loss estimates, typically through management overlays. We use LLM-generated forecasts as structured expert views to form adjustment signals, and incorporate them into the baseline using a Bayesian shrinkage framework (Stein, 1956; Jorion, 1986).



# Bayesian Shrinkage: Augmenting the Baseline



How it works:

- We start with a baseline forecast.
- The LLM generates multiple forecasts for each bank, and we take their average as a consensus view.
- The final estimate adjusts the baseline toward the LLM view, depending on how reliable the LLM signal is.

The framework adjusts itself:

- When LLM forecasts are inconsistent → the adjustment is small, and the estimate stays close to the baseline.
- When the baseline result is unreliable → the estimate shifts more toward the LLM.
- When more LLM forecasts are used to get the consensus view → the LLM signal becomes more reliable and the estimate shifts more toward the LLM view.

# Results: Pre-Trained LLM with Bayesian Shrinkage



We evaluated the Bayesian shrinkage framework using a pre-trained LLM (Qwen3-1.7B) across one-year to five-year prediction horizons, and compared it against the econometric baselines <sup>(Harris et al., 2018; Lu & Nikolaev, 2022)</sup> and the regression-based augmented model.

Forecast accuracy:

- The Bayesian augmented model achieves at least 37.7% MSE and 41.6% MAE improvements over the baseline across all horizons, with the max MSE and MAE improvements of 74.2% and 61.6% at five-year prediction horizons.

Bayesian vs. regression-based augmentation:

- The Bayesian approach outperforms the regression-based approach across all horizons, with at least 23.5% MSE and 31.5% MAE improvements over the regression model, because it assigns bank-specific weights rather than a single weight at each time point.

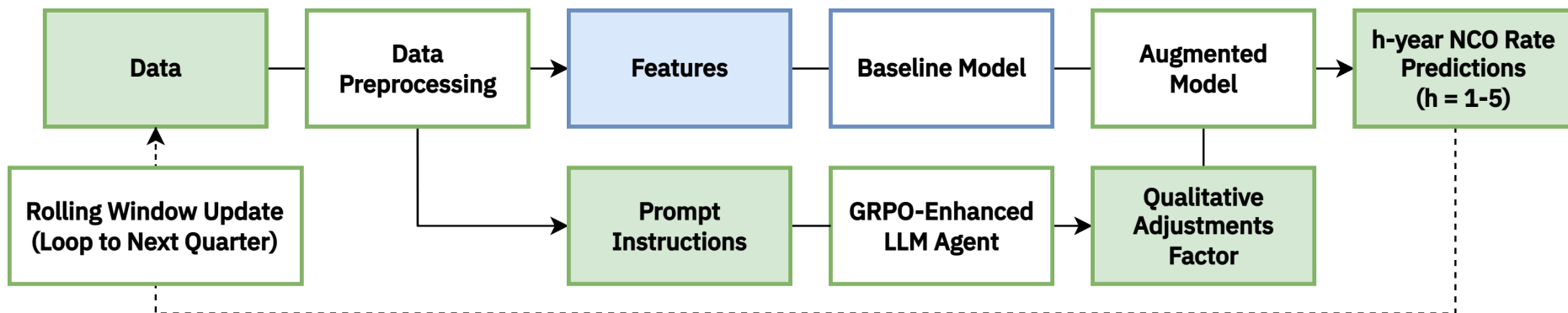
Key takeaway: Even with a pre-trained general-purpose LLM, the Bayesian framework extracts useful signal and improves baseline forecasts, establishing the value of the augmentation approach.

# GRPO: Fine-Tuning the LLM Expert



**Why fine-tune?** Pre-trained LLMs are general-purpose. To serve as reliable expert views in the Bayesian framework, the LLM needs to reason specifically about credit loss estimation.

How it works: GRPO (Group Relative Policy Optimization; Shao et al., 2024) fine-tunes the LLM by generating a group of candidate forecasts for each bank and each quarter, scoring them against actual outcomes via a reward function, and updating the model to favor better candidates over weaker ones within the same group.



# Predictors for NCO Rate



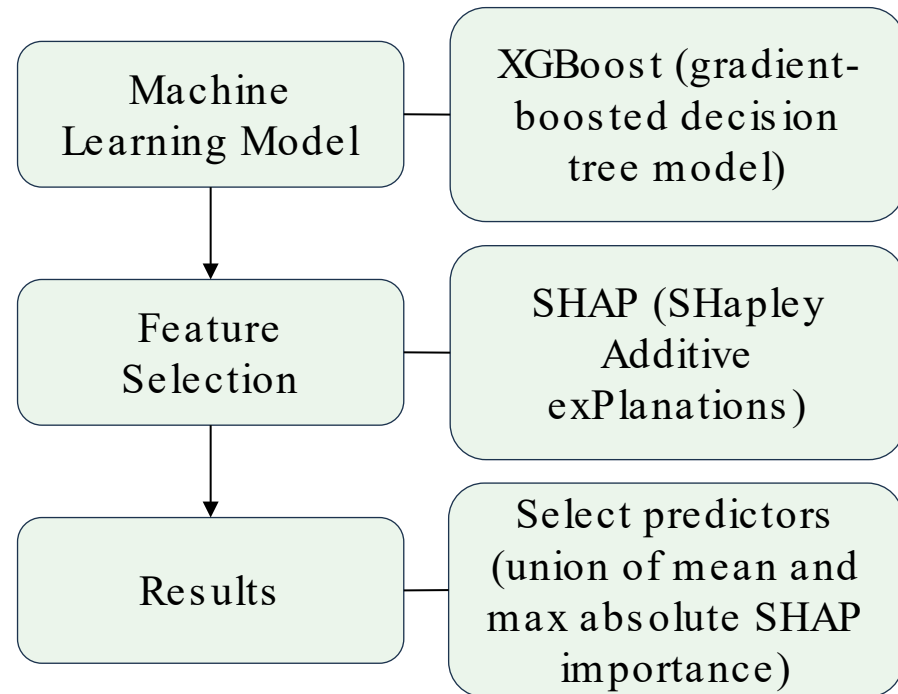
- Number of Banks: 1,196
- Sample Period: 2001 Q1 – 2025 Q2

| Predictor Type          | Definition   | Source                               | Examples   |
|-------------------------|--|--------------------------------------|--|
| Bank-level variables    | Internal bank-specific credit risk indicators (Lu & Nikolaev, 2022; Harris et al. 2018)                | FR Y-9C reports                      | Current NCO rate, Nonperforming loans, Average loan yields, Floating-rate loan share |
| Macroeconomic variables | Economy-wide conditions affecting borrower repayment ability (Lu & Nikolaev, 2022; Harris et al. 2018) | FRED-QD; Bureau of Economic Analysis | Unemployment rate, Inflation rate, GDP growth, CPI, HPI, and Disposable income       |

# Feature Selection Process



- Objective: Select a subset of bank-level and macroeconomic variables that will be added in the GRPO-LLM prompt.
- Rationale: The full set of variables (245 variables) cannot be incorporated due to prompt context length constraints (~1500 tokens).





# GRPO: Initialization (Model and Environment)

- Load Language Model: Pre-trained language model (Qwen3-1.7B).
- Set Learning Rules (Hyperparameters):
  - Learning speed
  - Penalizes large deviations from prior updates
  - Degree of exploration in predictions
  - Number of questions to ask per iteration
- Build RL Environment:
  - Prompts: Feed bank and macro data → ask "What's next quarter's NCO rate?"
  - Reward: Measure prediction accuracy against actual outcomes, confidence level, prediction stability, and reasoning.

| Prompt Section (What it is)   | Brief Explanation (Why it matters)   |
|---|--|
| <b>1. ROLE</b><br>You are a senior credit risk analyst.....   | Sets the role for the LLM so it responds from the perspective of an expert.  |
| <b>2. TARGET &amp; TIME HORIZON</b><br><b>TARGET:</b> Forecast the NCO rate 1 year(s) ahead for Bank with ID 1103766.<br><b>Input Quarter:</b> 2009Q3 → <b>Output Quarter:</b> 2010Q3   | Defines what to predict, for which bank, and the forecast horizon.   |
| <b>3. BANK-LEVEL FEATURES</b><br>(at Input Quarter 2009Q3): <ul style="list-style-type: none"> <li>– NCO rate at input quarter (annualized): 0.1014%</li> <li>– YoY change in unexpected NPL ratio: -0.03219 pp</li> <li>– NPL ratio (of total loans): 0.3383%</li> <li>– Average loan yield (annualized): 12.89%</li> <li>– Floating-rate loan share: 31.5%</li> <li>– Real estate loan share: 71.51%</li> <li>– Consumer loan share: 8.488%</li> </ul>  | Provides bank-specific financial and risk metrics at the starting time. These describe the bank's current condition. |
| <b>4. MACROECONOMIC FEATURES</b><br>(Note: 'growth' values are quarterly log-differences in decimal form, e.g., 0.02 means 2% growth) <ul style="list-style-type: none"> <li>– U. Michigan consumer sentiment (index) [current qtr]: 68.4</li> <li>– U. Michigan consumer sentiment (index) [1-qtr lag]: 68.2</li> <li>– Financial sector employment growth [2-qtr lag]: -0.0163</li> <li>– Real household real estate assets growth [2-qtr lag]: -0.02864</li> <li>– Construction employment growth [1-qtr lag]: -0.05538</li> <li>– ...</li> </ul>  | Adds macroeconomic context that can affect future credit losses.   |
| <b>5. FORECASTING TASK</b><br>Using features observed at quarter t (2009Q3), forecast the bank's NCO rate at t+1Y (2010Q3).   | Clearly states the forecasting objective and time alignment.   |
| <b>6. REQUIREMENTS</b> <ol style="list-style-type: none"> <li>1. Incorporate bank-level risk indicators (portfolio mix, credit quality, pricing, rate sensitivity)</li> <li>2. Incorporate macroeconomic conditions</li> <li>3. Explicitly reason about:               <ol style="list-style-type: none"> <li>(a) Macro-driven loss rate shifts over the next 1 year(s)</li> <li>(b) Bank-specific amplification or mitigation factors</li> </ol> </li> <li>4. Quantify uncertainty via prediction intervals</li> <li>5. Keep your analysis concise and focused — aim for 450-600 words total</li> </ol>                  | Sets analytical expectations: what to consider, what to explain, and how to handle uncertainty and length.           |
| <b>7. OUTPUT FORMAT (STRICTLY FOLLOW)</b><br><b>&lt;analysis&gt;</b><br>Step 1: Analyze bank credit quality, portfolio composition, and rate sensitivity<br>Step 2: Analyze macroeconomic impact on expected losses<br>Step 3: Synthesize bank and macro factors to justify NCO trajectory<br>Step 4: Assess uncertainty sources and prediction interval width<br><b>&lt;/analysis&gt;</b><br><b>&lt;prediction&gt;</b><br>NCO_Rate_At_2010Q3: [value]%<br>Prediction_Interval: [lower]%, [upper]%<br>Primary_Drivers: [comma-separated key drivers]<br>Confidence_Level: [high/medium/low]<br><b>&lt;/prediction&gt;</b> | Specifies the exact structure of the response for easy parsing and consistency.                                      |
| <b>8. FINAL INSTRUCTION</b><br>All output values in PERCENTAGE.<br>Begin your analysis and forecasts:   | Reminds the model of units and signals the start of the response.  |

### Example Prompt

You are a senior credit risk analyst.....

**TARGET:** Forecast the NCO rate 1 year(s) ahead for Bank with ID 1103766.

**Input Quarter:** 2009Q3 → **Output Quarter:** 2010Q3

#### **BANK-LEVEL FEATURES (at Input Quarter 2009Q3):**

- NCO rate at input quarter (annualized): 0.1014%
- YoY change in unexpected NPL ratio: -0.03219 pp
- NPL ratio (of total loans): 0.3383%
- Average loan yield (annualized): 12.89%
- Floating-rate loan share: 31.5%
- Real estate loan share: 71.51%
- Consumer loan share: 8.488%

#### **MACROECONOMIC FEATURES:**

(Note: 'growth' values are quarterly log-differences in decimal form, e.g., 0.02 means 2% growth)

- U. Michigan consumer sentiment (index) [current qtr]: 68.4
- U. Michigan consumer sentiment (index) [1-qtr lag]: 68.2
- Financial sector employment growth [2-qtr lag]: -0.0163
- Real household real estate assets growth [2-qtr lag]: -0.02864
- Construction employment growth [1-qtr lag]: -0.05538
- ...

#### **FORECASTING TASK:**

Using features observed at quarter t (2009Q3), forecast the bank's NCO rate at t+1Y (2010Q3).

#### **Requirements:**

1. Incorporate bank-level risk indicators (portfolio mix, credit quality, pricing, rate sensitivity)
2. Incorporate macroeconomic conditions
3. Explicitly reason about:
  - (a) Macro-driven loss rate shifts over the next 1 year(s)
  - (b) Bank-specific amplification or mitigation factors
4. Quantify uncertainty via prediction intervals
5. Keep your analysis concise and focused — aim for 450-600 words total

#### **OUTPUT FORMAT (STRICTLY FOLLOW):**

##### **<analysis>**

Step 1: Analyze bank credit quality, portfolio composition, and rate sensitivity

Step 2: Analyze macroeconomic impact on expected losses

Step 3: Synthesize bank and macro factors to justify NCO trajectory

Step 4: Assess uncertainty sources and prediction interval width

##### **</analysis>**

##### **<prediction>**

NCO\_Rate\_At\_2010Q3: [value]%

Prediction\_Interval: [lower]%, [upper]%

Primary\_Drivers: [comma-separated key drivers]

Confidence\_Level: [high/medium/low]

##### **</prediction>**

**All output values in PERCENTAGE. Begin your analysis and forecasts:**

### Example Response

#### **<analysis>**

##### **Step 1: Analyze bank credit quality, portfolio composition, and rate sensitivity**

The bank has a relatively low NCO rate of 0.1014% at the input quarter (2009Q3), indicating a strong credit quality. The bank's portfolio is composed of a significant share of consumer loans (8.488%), real estate loans (71.51%), and floating-rate loans (31.5%), which suggests a diversified risk profile. The average loan yield is 12.89%, which is relatively high, indicating a conservative pricing strategy.

The bank's willingness to lend to consumers is -5.9 at the 1-qtr lag, suggesting a reduced lending stance to consumers, which could lead to increased NCO rates. The real estate loan share is high, and the real estate sector is currently experiencing a slowdown in growth, which could lead to increased NCO rates. The floating-rate loan share is also high, and the current interest rate environment is relatively stable, which could lead to reduced NCO rates.

##### **Step 2: Analyze macroeconomic impact on expected losses**

The macroeconomic environment in the past year is mixed. The U. Michigan consumer sentiment is 68.4, which is above the 68.2 from the previous quarter, suggesting a slight improvement in consumer confidence. This could lead to reduced NCO rates.

The Fed Funds rate is -0.0233%, which is slightly negative, indicating a slight decrease in the interest rate. This could lead to reduced NCO rates. The inventory investment share of GDP is -1.4%, which is negative, indicating a slight decrease in inventory levels. This could lead to reduced NCO rates.

The Aaa-Fed Funds spread is 5.11%, which is relatively high, indicating a slight increase in the cost of borrowing. This could lead to reduced NCO rates. The real nonrevolving consumer credit growth is -0.0061244%, which is negative, indicating a slight decline in consumer credit availability. This could lead to reduced NCO rates.

##### **Step 3: Synthesize bank and macro factors to justify NCO trajectory**

The bank's low NCO rate and diversified portfolio suggest a relatively low risk of charge-offs. However, the macroeconomic environment in the past year is mixed, with some indicators suggesting a slight improvement in the economy and others suggesting a slight increase in the cost of borrowing.

The bank's willingness to lend to consumers is reduced, which could lead to increased NCO rates. The real estate loan share is high, and the real estate sector is currently experiencing a slight slowdown in growth, which could lead to reduced NCO rates. The floating-rate loan share is high, and the current interest rate environment is relatively stable, which could lead to reduced NCO rates.

##### **Step 4: Assess uncertainty sources and prediction interval width**

The uncertainty sources include the macroeconomic environment, the bank's credit quality, and the bank's portfolio composition. The prediction interval is likely to be relatively wide, as the macroeconomic environment is uncertain and the bank's credit quality is relatively stable.

##### **</analysis>**

##### **<prediction>**

NCO\_Rate\_At\_2010Q3: 0.16%

Prediction\_Interval: 0.00%, 1.20%

Primary\_Drivers: NCO\_rate, macroeconomic\_factors, bank\_portfolio\_composition, bank\_credit\_quality

Confidence\_Level: Medium

##### **</prediction>**

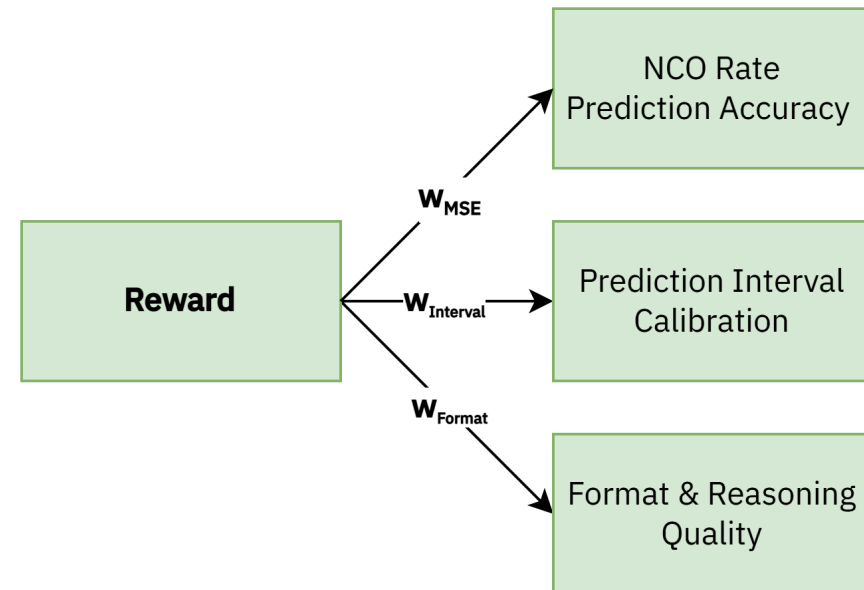
# GRPO - Fine-Tuning Iteration



## Phase 1: GRPO Fine-Tuning Data

### Preparation (per Round)

1. Provide the model with the prompt containing bank-specific content.
2. Generate multiple responses (e.g., 8) with slight randomness.
3. Evaluate each response using the reward function.
4. Compare responses with the group and determine relative quality (standardize scores).



# Reward Components

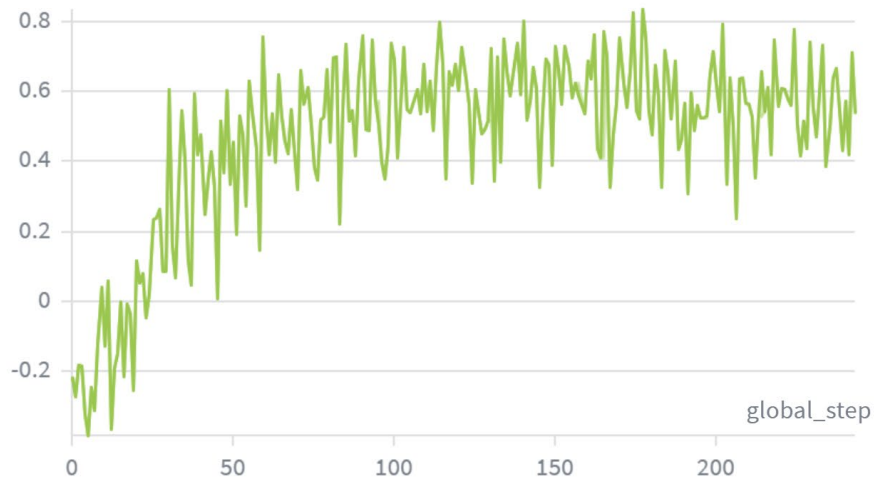


| Reward Component                 | Objective  | Weight |
|----------------------------------|--|--------|
| Prediction Accuracy              | Measures numeric accuracy of the next-quarter NCO prediction                     | 0.8    |
| Confidence Interval Calibration  | Rewards well-calibrated and accurate confidence intervals of NCO rate prediction | 0.1    |
| Reasoning Quality and Formatting | Measures how thoroughly the model explains its prediction                        | 0.1    |

# Reward Convergence during GRPO Fine-Tuning



Train/Interval Reward



Train/Format Reward



# GRPO: Fine-Tuning Iteration



## Phase 2: Model Update (Per Data Chunk/Round)

1. Compute Loss Function (Optimization Objective Function)
  1. Prioritize better responses while limiting large changes to the model.
  2. Adjust the model not to deviate significantly from previous policy.
  3. Add entropy term to maintain diversity and avoid premature convergence.
2. Update Model Parameters
  1. Adjust parameters using gradient descent to reduce loss.
  2. Apply clipping and small updates to ensure stability.

# GRPO: Fine-Tuning Iteration



## Phase 3: Evaluation and Checkpoints (per Round)

1. Run periodic tests on out-of-sample validation data to detect overfitting or poor generalization.
2. Save checkpoints (model parameters + full metadata: model version, iteration number, timestamp, key hyperparameters, progress metrics) every N steps for reproducibility.

# Results: Fine-Tuned LLM with Bayesian Shrinkage



We fine-tuned the LLM using GRPO and compared its performance against the pre-trained LLM within the Bayesian framework. Results are currently available for one-year and two-year horizons.

Fine-tuned vs. pre-trained (within the Bayesian framework):

- One-year horizon: the Bayesian model using fine-tuned LLM forecasts achieves 23.4% MSE and 18.4% MAE improvement over the Bayesian model using pre-trained forecasts.
- Two-year horizon: the improvement remains positive at 9.6% MSE and 21.8% MAE.

Key takeaway:

- Fine-tuning produces more accurate expert views, which the Bayesian framework rewards with greater weight —further improving the augmented forecast. Longer-horizon results are in progress.

Updating fine-tuning setup:

- We are refining the GRPO framework by using a binary reward (threshold-based) to improve training stability, aligning with standard GRPO settings that emphasize binary accuracy signals.

# Results: Robustness and Look-Ahead Bias



Robustness test with a different pre-trained LLM:

- Besides Qwen3-1.7B, we tested the Bayesian framework with another pre-trained LLM (Qwen2.5-3B-Instruct) to check whether the results depend on a specific model.
- Across both LLMs, the Bayesian model consistently delivers strong gains over both baseline and regression models across all prediction horizons:
  - MSE improvement over baseline:  $\sim 30\text{--}74\%$  (Bayesian) vs.  $\sim 1\text{--}32\%$  (regression)
  - MAE improvement over baseline:  $\sim 34\text{--}62\%$  (Bayesian) vs.  $\sim 0.8\text{--}24\%$  (regression)

Look-ahead bias:

- We split the sample into two periods: before and after each LLM's knowledge cutoff date (Qwen3-1.7B: 2024Q4; Qwen2.5-3B: 2024Q1), to test whether the LLM is relying on information it should not have access to at prediction time.
- Performance is consistent before and after the cutoff:  
Bayesian improvements remain high —MSE:  $\sim 70\text{--}74\% \rightarrow \sim 73\text{--}79\%$ ; MAE:  $\sim 55\text{--}62\% \rightarrow \sim 66\text{--}68\%$ .
- Regression improvements are also stable before and after the cutoff:  
MSE:  $\sim 2\text{--}3\% \rightarrow \sim 4\text{--}7\%$ ; MAE:  $\sim 1\text{--}2\% \rightarrow \sim 2\text{--}4\%$ .

# Conclusion



- Developed a Bayesian framework that combines econometric baselines with LLM forecasts for CECL estimation
  - Pre-trained LLM (Qwen3-1.7B, Bayesian model): at least 37.7% MSE and 41.6% MAE improvement over baseline; up to 74.2% MSE and 61.6% MAE (5-year)
  - Outperforms regression-based augmentation by at least 23.5% MSE and 31.5% MAE (due to bank-specific weighting)
- Robust across pre-trained models (Qwen3-1.7B, Qwen2.5-3B): Bayesian MSE ~29–74% and MAE ~34–62% across horizons → gains come from the framework, not a specific model
- No look-ahead bias: performance remains stable before/after cutoff (Bayesian MSE ~70–74% → ~73–79%, MAE ~55–62% → ~66–68%)
- Applying the GRPO approach to fine-tune an LLM in order to further improve Bayesian results

# Takeaways & Next Steps



## Takeaways

- AI has the potential to augment the accounting estimation process, with human expertise remaining essential for guiding model design and validating outputs.
- Designing effective reward functions is iterative, requiring human judgment to balance objectives (accuracy, interval quality, and response format consistency) and diagnose training failures.

## Next Steps

- Increase the number of LLM forecasts per bank (from 5 to 30) to strengthen the consensus signal.
- Extend the Bayesian shrinkage framework to relative views across banks (e.g., how does one bank compare to another?), connecting to peer assessment in CECL practice.
- Test with additional pre-trained LLMs (e.g., Llama, Gemma) to evaluate robustness.
- Continue to improve the GRPO framework and fine-tune the LLM.

# Thank You!





# XBRL Enhanced LLM Agents for Financial Analytics

AI and Structured Data Forum: Optimizing Performance

Dr. Steve Yang

School of Business  
Stevens Institute of Technology

May 15, 2026

# Outline

Motivation: Why XBRL Enhanced LLMs

Methodology: De-noise with RAG and FinLoRA using XBRL

Results: Performance Gains from the XBRL Enhanced LLMs

Next Frontier: Accounting Knowledge Graph RAG Method

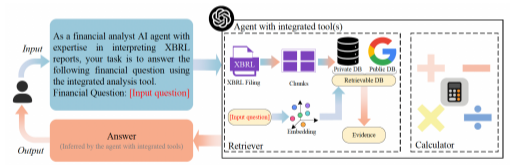
Conclusion: Toward Intelligent Financial Systems

# Why Do We Need XBRL-Enhanced LLMs for Financial Analysis?

- ▶ Financial analysis:
  - ▶ Compute valuation metrics,
  - ▶ Compare trends,
  - ▶ Normalize financial data,
  - ▶ Project future cash flows.
  
- ▶ Traditional LLMs struggle with:
  - ▶ Structured financial data
  - ▶ XBRL taxonomy understanding
  - ▶ Numerical reasoning
  - ▶ Financial formula construction
  
- ▶ XBRL provides:
  - ▶ Standardized semantics
  - ▶ Structured relationships
  - ▶ Machine-readable financial context

**Key Research Question**

Can LLMs leverage XBRL structure to become reliable financial analysis agents?



Han et al. (2024), *XBRL Agent: Leveraging Large Language Models for Financial Report Analysis*

# Can LLMs Understand XBRL?

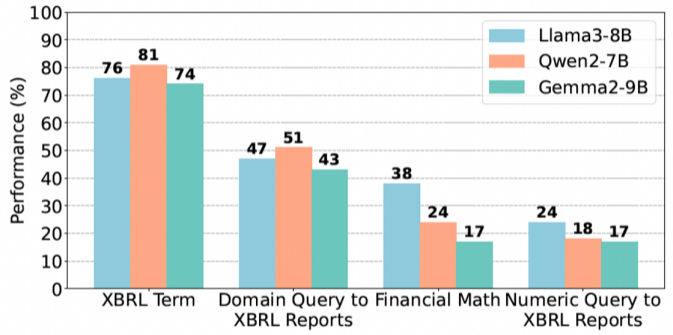
ICAIF '24, November 14–17, 2024, Brooklyn, NY, USA

Shijie Han, Haoqiang Kang, Bo Jin, Xiao-Yang Liu, and Steve Y Yang

| XBRL Term  | Domain Query to XBRL Reports  | Financial Math   | Numeric Query to XBRL Reports   |
|--|---|--|---|
| <p><b>Q:</b> What does the term <b>'abstract'</b> mean in the context of the XBRL standard? Please provide a detailed explanation of this term.</p> <p><b>A:</b> An attribute of an element to indicate that the element is only used in a hierarchy to group related elements together. An abstract element cannot be used to tag data in an instance document.</p> | <p><b>Q:</b> Among operations, investing, and financing activities, which brought in the most (or lost the least) cash flow for Nike in FY2023?</p> <p><b>A:</b> Among the three, cash flow from operations was the highest for Nike in FY2023.</p>   | <p><b>Q:</b> A project expects annual cash inflows of \$6,000 for 4 years. If the discount rate is 8%, what is the Net Present Value (NPV) of the project?</p> <p><b>A:</b> 21462.58</p> | <p><b>Q:</b> What is the FY2015 unadjusted EBITDA margin for Netflix? Calculate unadjusted EBITDA using unadjusted operating income and D&amp;A.</p> <p><b>A:</b> 0.054</p> |
| <p><b>Q:</b> What does the term <b>'fact'</b> mean in the context of the XBRL standard? Please provide a detailed explanation of this term.</p> <p><b>A:</b> The occurrence in an instance document of a value or other information tagged by a taxonomy element.</p>  | <p><b>Q:</b> Is 3M a capital-intensive business based on FY2022 data?</p> <p><b>A:</b> No, the company is managing its CAPEX and Fixed Assets pretty efficiently, which is evident from below key metrics: CAPEX/Revenue Ratio: 5.1%; Fixed assets/Total Assets: 20%; Return on Assets= 12.4%</p> | <p><b>Q:</b> Suppose you anticipate receiving \$10,000 in 5 years. If the annual discount rate is 7%, what is the present value of this sum?</p> <p><b>A:</b> 7129.86</p>                | <p><b>Q:</b> What is Amazon's FY2019 net income attributable to shareholders (in USD millions)?</p> <p><b>A:</b> 0.308</p>  |
| <p><b>Q:</b> What does the term <b>'label'</b> mean in the context of the XBRL standard? Please provide a detailed explanation of this term.</p> <p><b>A:</b> Human-readable name for an element; each element has a standard label that corresponds to the element name, and is unique across the taxonomy</p>  | <p><b>Q:</b> What industry does AMCOR primarily operate in?</p> <p><b>A:</b> Amcor is a global leader in packaging production for various use cases.</p>  | <p><b>Q:</b> An annual savings account has a nominal interest rate of 6%. If interest is compounded quarterly, what is the Effective Annual Rate (EAR)</p> <p><b>A:</b> 0.06</p>         | <p><b>Q:</b> What is the year end FY2019 total amount of inventories for Best Buy? Answer in USD millions</p> <p><b>A:</b> 5409</p>   |
| <p>Support material (Optional provided): Public official glossary of XBRL terminology</p>  | <p>Support material (Optional provided): XBRL reports of related companies</p>  | <p>Support material (Optional provided): Private financial formulas explanation database</p>   | <p>Support material (Optional provided): XBRL reports of related companies</p>  |

# Motivating Results: LLM Limitations

- ▶ Moderate performance on XBRL terminology.
- ▶ Weak performance on domain-specific report queries.
- ▶ Poor performance on financial math and numeric reasoning.



# Key Limitations of Pure LLMs

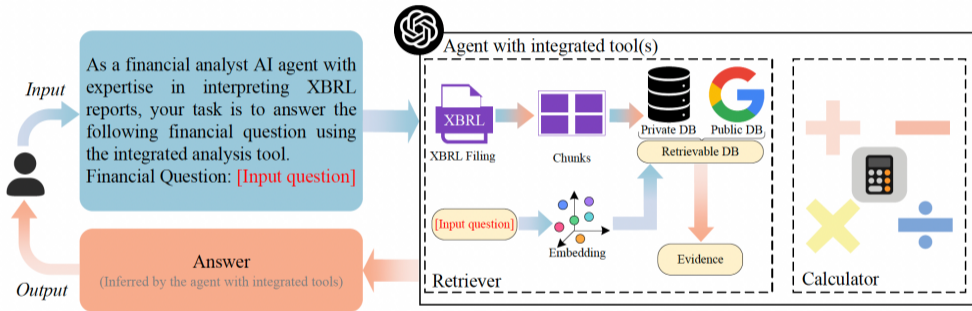
- ▶ Limited financial domain knowledge
- ▶ Weak mathematical reasoning
- ▶ Difficulty in interpreting structured XBRL data

**Implication:** pure LLMs are not sufficient for reliable XBRL-based financial analysis.

Table 1: Case study—XBRL tagging (Google 10-Q 2025-Q1) and XBRL formula calculation (Travelers 10-K FY-2023)—from base Llama 3.1 8B Instruct and our LoRA fine-tuned version.

|                             |   |
|-----------------------------|---|
| <b>Question 1</b>           | What is the appropriate XBRL US-GAAP tag for “2.0” in 3 “...equity securities accounted for under the equity method had a carrying value of approximately \$2.0 billion”? |
| Llama 3.1 8B                | us-gaap:MajorityEquityInterest  |
| Llama 3.1 8B LoRA (8bit r8) | us-gaap:EquityMethodInvestments   |
| Ground truth                | us-gaap:EquityMethodInvestments   |
| <b>Question 2</b>           | What is Travelers Companies Inc’s <b>Equity Multiplier</b> for FY 2023? (Answer with a formula substituted with values.) {XBRL Context}                                   |
| Llama 3.1 8B                | (1,209,000,000 / 249,210,000,000)   |
| Llama 3.1 8B LoRA (8bit r8) | 125,978,000,000 / 249,210,000,000   |
| Ground truth                | 125,978,000,000 / 249,210,000,000   |

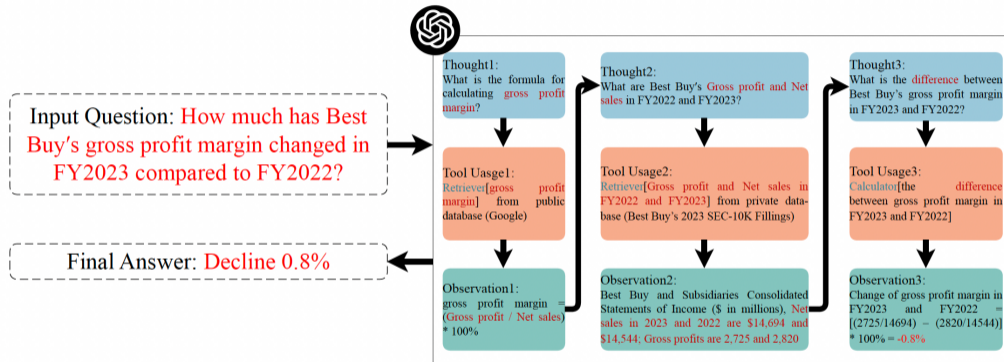
# Toward XBRL Financial Analysis Agents



# Agent Reasoning Workflow

ICAIF '24, November 14–17, 2024, Brooklyn, NY, USA

Shijie Han, Haoqiang Kang, Bo Jin, Xiao-Yang Liu, and Steve Y Yang



## Core Idea

The agent combines retrieval, calculation, and reasoning to answer financial questions.

# Methodology: RAG and FinLoRA with XBRL

► **Problem:**

- LLMs lack financial/XBRL knowledge

► **Solution: FinLoRA**

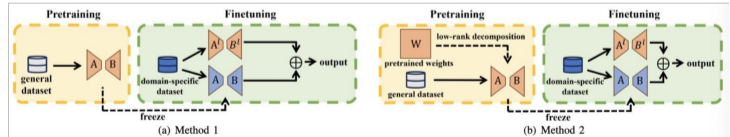
- Parameter-efficient fine-tuning
- Train on financial + XBRL datasets

► **Core idea:**

Learn domain-specific representations via low-rank updates

► **Key advantages:**

- Gained significant improvements in "Domain Query to XBRL Reports" and "Financial Math" performance
- But still low in numeric reports



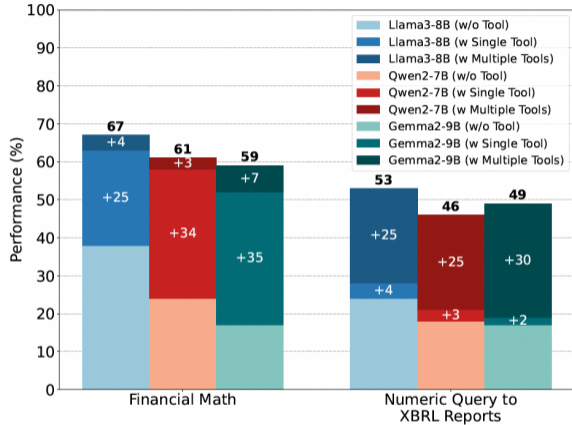
# Why Fine-Tuning is Needed

- ▶ General LLMs fail on financial tasks:
  - ▶ Misinterpret XBRL tags
  - ▶ Incorrect financial calculations
- ▶ Example:

Base model selects wrong XBRL tag; LoRA fixes it

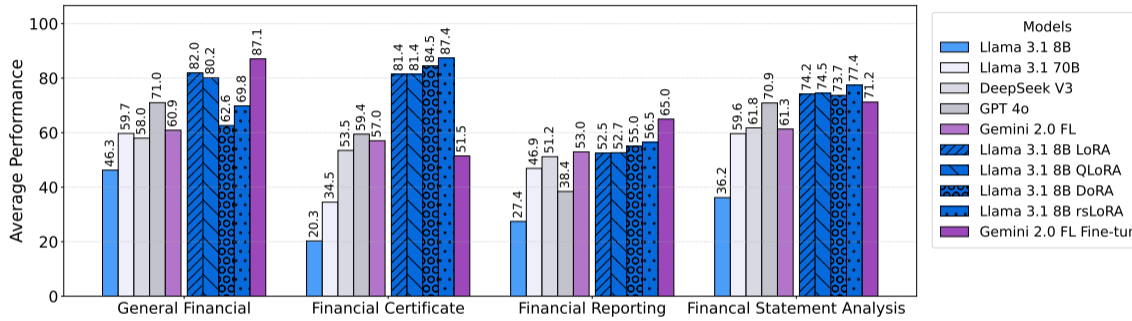
- ▶ Root causes:
  - ▶ Lack of financial data in pretraining
  - ▶ Complex structured formats (XBRL)

# LLM Fintuning Performance Gains



- ▶ Average improvement: **+25/+35%** for financial math, and **+25/+30%** on XBRL numbers.

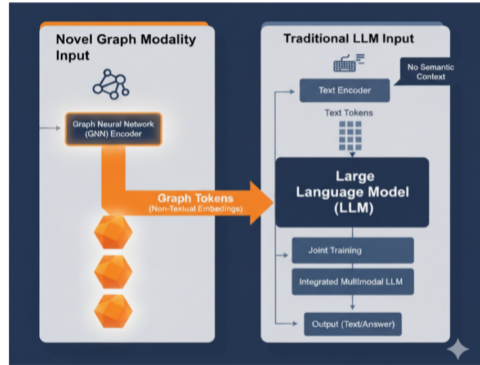
# FinLoRA Performance Gains



- ▶ LoRA-based models significantly outperform base LLMs
- ▶ Average improvement: **+36%**

# Next Step: Knowledge GraphRAG Finetuning

- ▶ **Limitation:** All information, regardless of its native structure, must be serialized or “linearized” into a one-dimensional sequence of text tokens.
- ▶ **Solution:** A Natively Graph-Aware LLM Architecture
  - ▶ The “Graph-as-a-Token? Modality
  - ▶ Semantically-Rich Retrieval using Graph Embeddings
  - ▶ Joint Multimodal Training



# Conclusion: Taking Advantage of the Full Potential of XBRL

| Companies  | Auditors   | Investors / information intermediaries   | Regulators   |
|--|--|--|--|
| <ul style="list-style-type: none"> <li>• Disclosure of high-quality XBRL data to help widen the discoverability of a company's capital offerings and reporting.</li> <li>• Fully integrate XBRL into the financial reporting process.</li> <li>• Lower the cost and increase the quality of internal or external reports.</li> </ul> | <ul style="list-style-type: none"> <li>• Streamline risk assessments and analytical procedures.</li> <li>• Automate audit tests using the tagged data from company information systems.</li> </ul> | <ul style="list-style-type: none"> <li>• Access information with greater speed and precision, all at a lower cost.</li> <li>• Enable deeper analysis with drill-down capability from the summary numbers</li> <li>• Comparisons of peer data and time series information.</li> </ul> | <ul style="list-style-type: none"> <li>• Enhanced regulatory monitoring, surveillance, and enforcement activities</li> <li>• Enhanced evidence to support public policy judgments.</li> <li>• Improving information accessibility and leveling the informational playing field among investors.</li> </ul> |

Source: XBRL US investor forum presentation

## Vision

Develop AI chatbots and agents capable of interpreting and answering questions using XBRL data.

# Conclusion: Toward Intelligent Financial Systems

## ▶ Key Insight

- ▶ XBRL provides structured, high-quality financial information and knowledge
- ▶ LLMs alone are insufficient for reliable financial analysis

## ▶ Our Framework

- ▶ **Agent Systems**: reasoning, retrieval, and tool integration
- ▶ **RAG + FinLoRA**: efficient domain adaptation for financial intelligence
- ▶ **GraphRAG**: understanding accounting domain knowledge for financial intelligence

## ▶ CRAFT Impact

- ▶ Industry-ready AI solutions for financial institutions
- ▶ Bridging structured data (XBRL) with generative AI
- ▶ Enabling transparency, efficiency, and innovation in capital markets



**STEVENS**

INSTITUTE OF TECHNOLOGY

1870

---

## School of Business

Steve Yang

syang14@stevens.edu

# Optimizing Performance



## Standards and Machine Learning

Xiao-Yang Liu Yanglet, Ph.D, Fu Foundation School of Engineering and Applied Science, Columbia University;  
Researcher, SecureFinAI LAB

# Fine-Tuning FinGPT for XBRL Using Low-Rank Adaptation

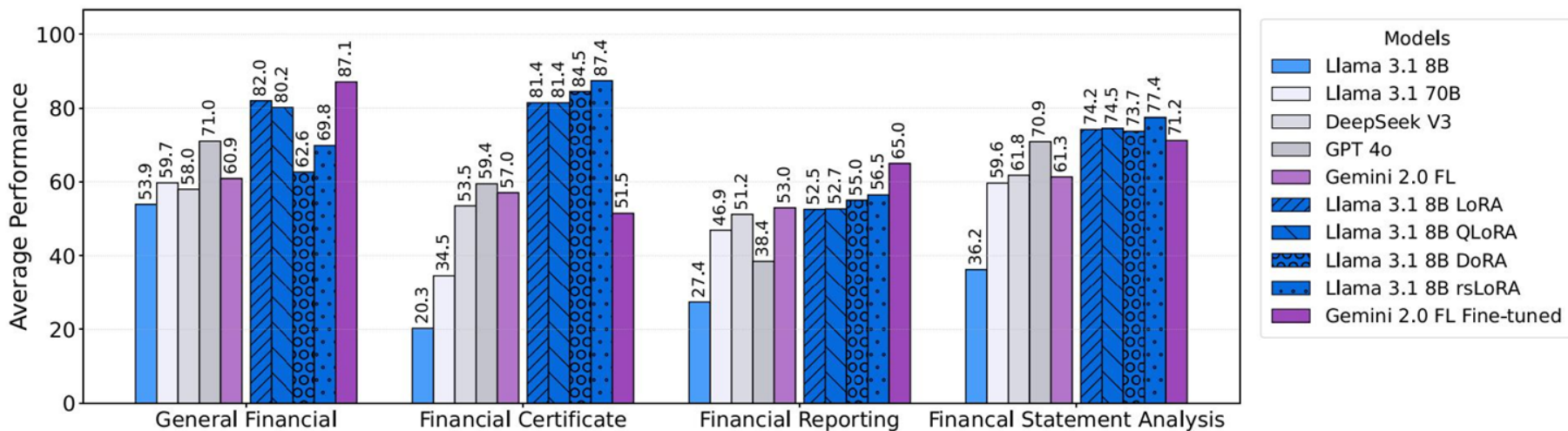
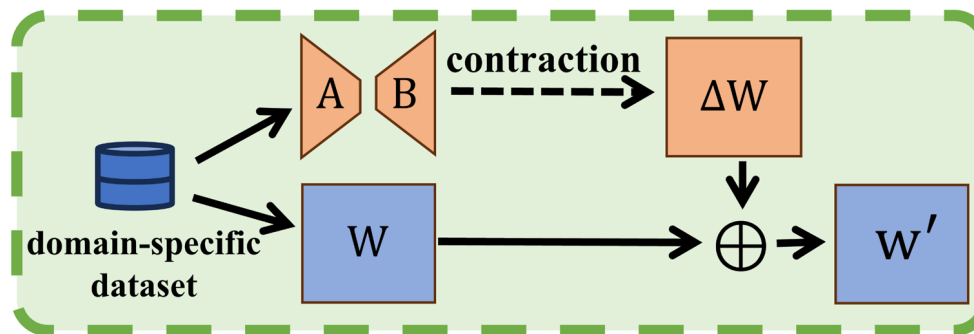


Figure 1: Average performance of base models and LoRA models.

# FinAuditing: Multi-Document Benchmark

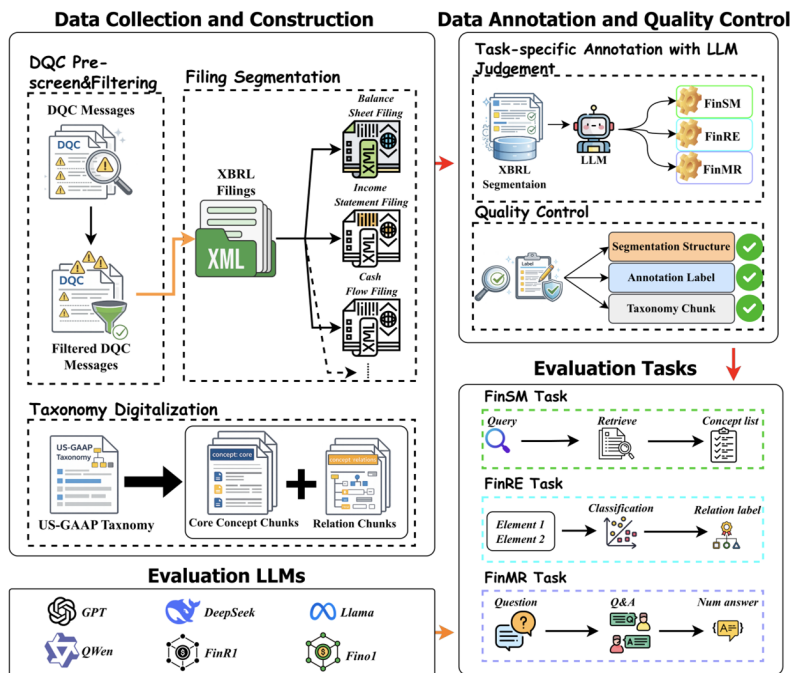


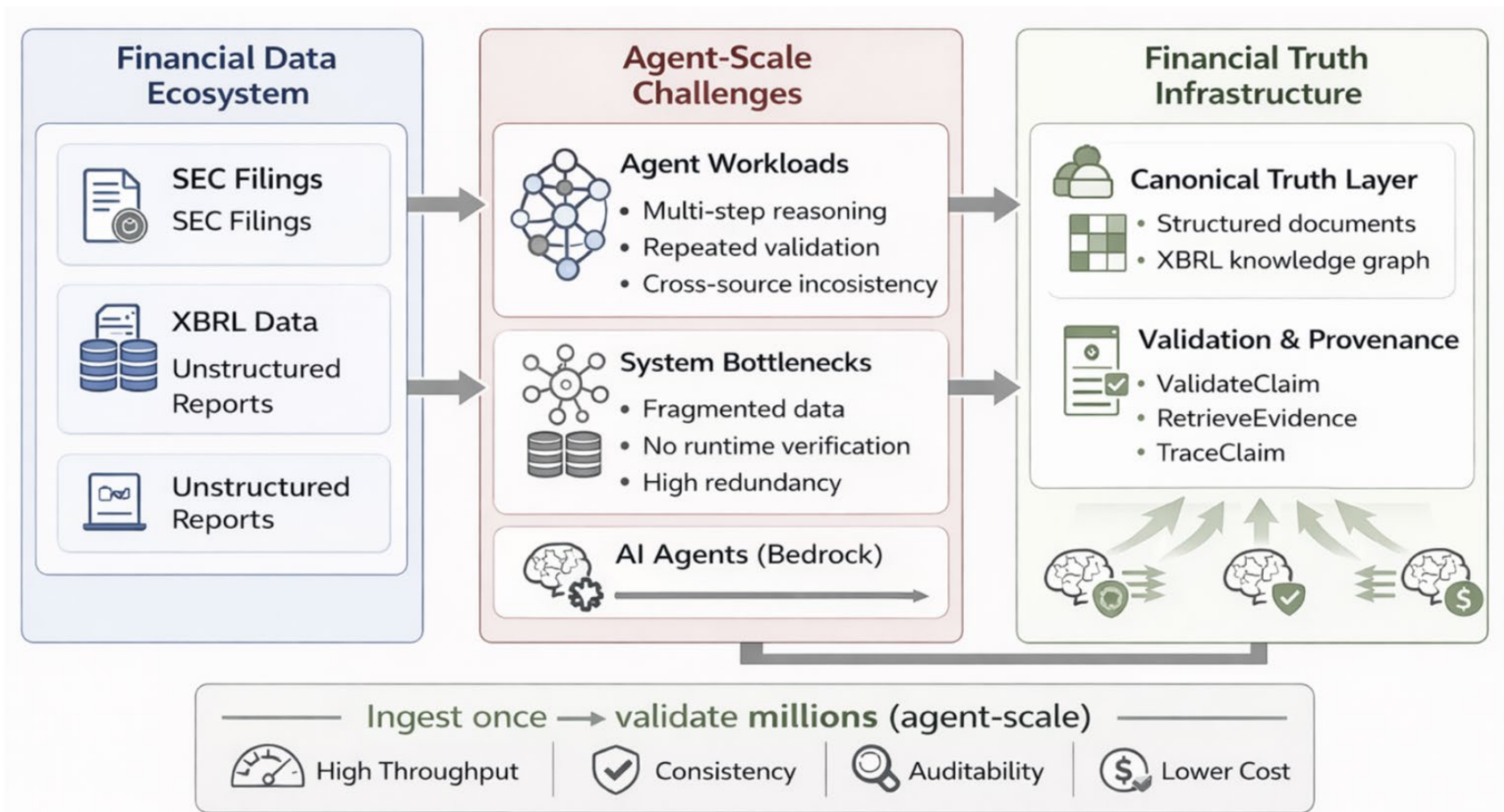
Figure 2: Overview of the **FINAUDITING** construction pipeline.

Table 2: Distribution of common error types in XBRL filings.

| Broad Category    | Error Type                                    | # Cases      | Proportion (%) | DQC US ID |
|-------------------|---|--------------|----------------|-----------|
| Semantic-based    | FS with no associated calculation             | 386          | 8.49           | DQC_0099  |
|                   | Concentration risk                            | 217          | 4.77           | DQC_0109  |
|                   | Location axis with a single member            | 208          | 4.58           | DQC_0137  |
| Relation-based    | Sibling and child relationships               | 570          | 12.54          | DQC_0081  |
|                   | Axis with inappropriate members               | 313          | 6.89           | DQC_0001  |
|                   | Inappropriate cash flow presentation          | 219          | 4.82           | DQC_0145  |
| Calculation-based | FS calculation check with no dimensional data | 260          | 5.72           | DQC_0126  |
|                   | Negative values                               | 368          | 8.10           | DQC_0015  |
|                   | FS tables dimensional cross check             | 202          | 4.44           | DQC_0117  |
| <b>Total</b>      |   | <b>2,743</b> | <b>60.33</b>   | -         |

# Financial Truth Infrastructure for AI Agents

## The Provenance-Validation-Compliance Layer for Agent-scale Systems

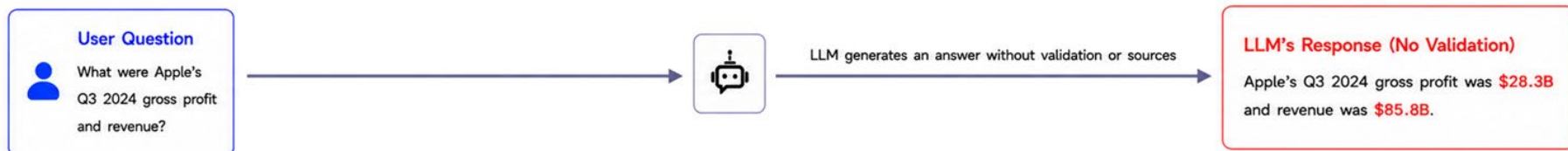


# Financial Truth Infrastructure for AI Agents

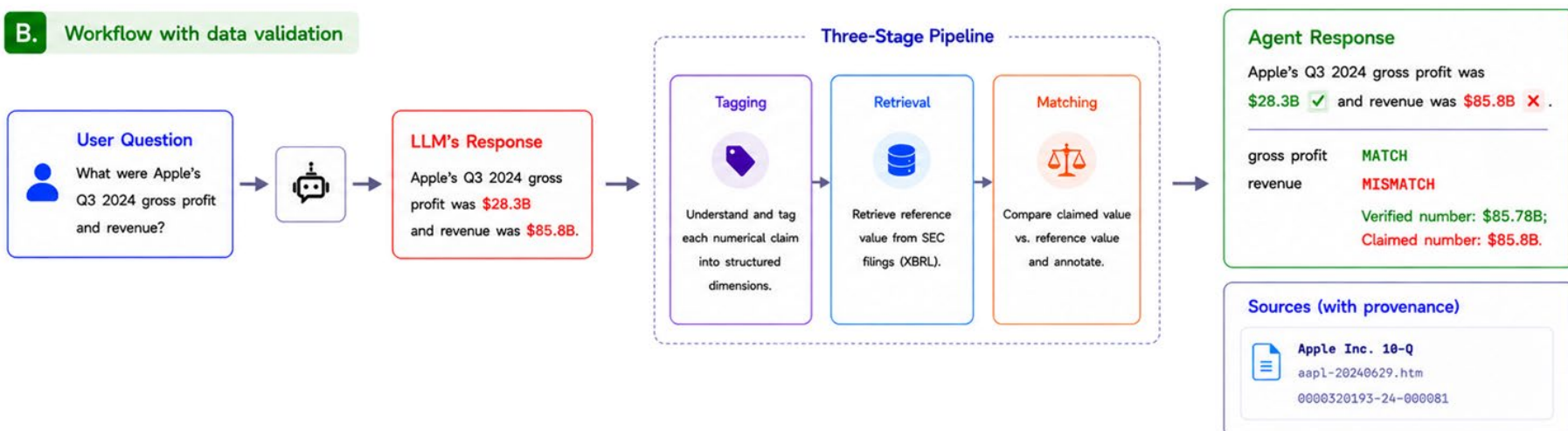
## Validation: Tagging-Retrieval-Matching Pipeline



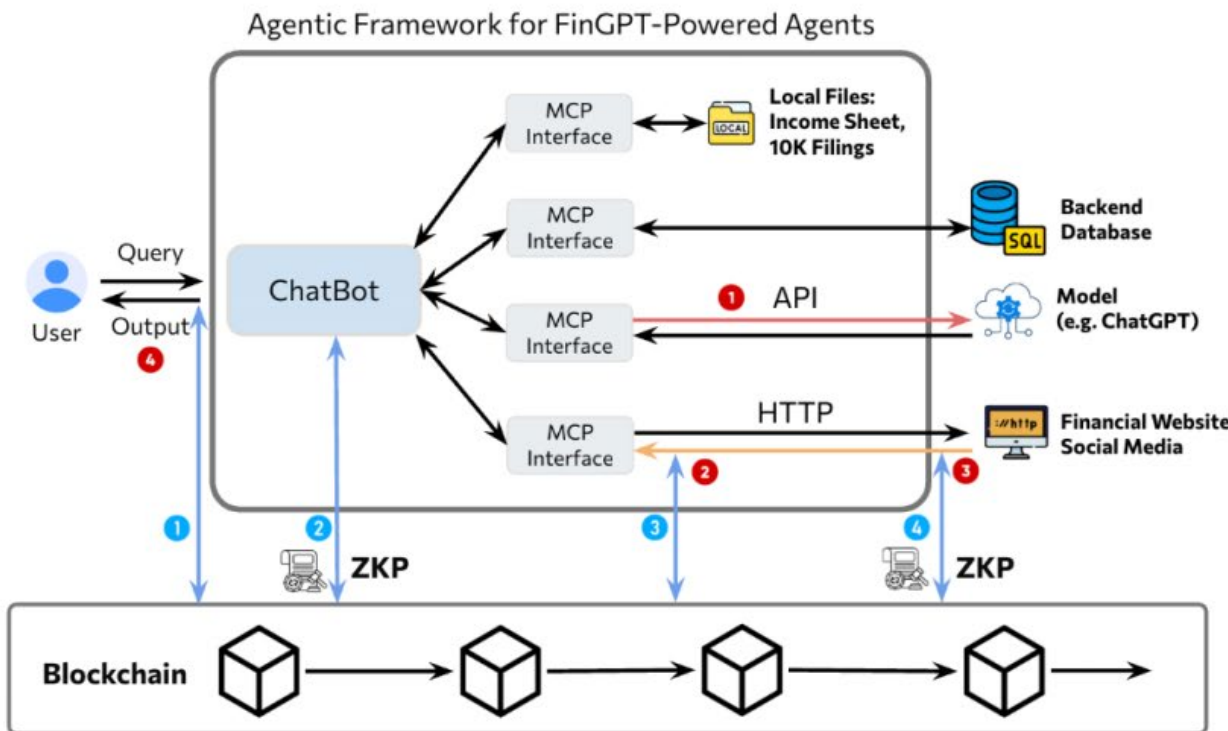
### A. Current workflow



### B. Workflow with data validation



# Guardrails for Secure and Privacy-Preserving FinAgents



Guardrail Framework