



Can AI Reasoning Improve Accounting Estimation? Evidence from CECL Allowance

Arion Cheong

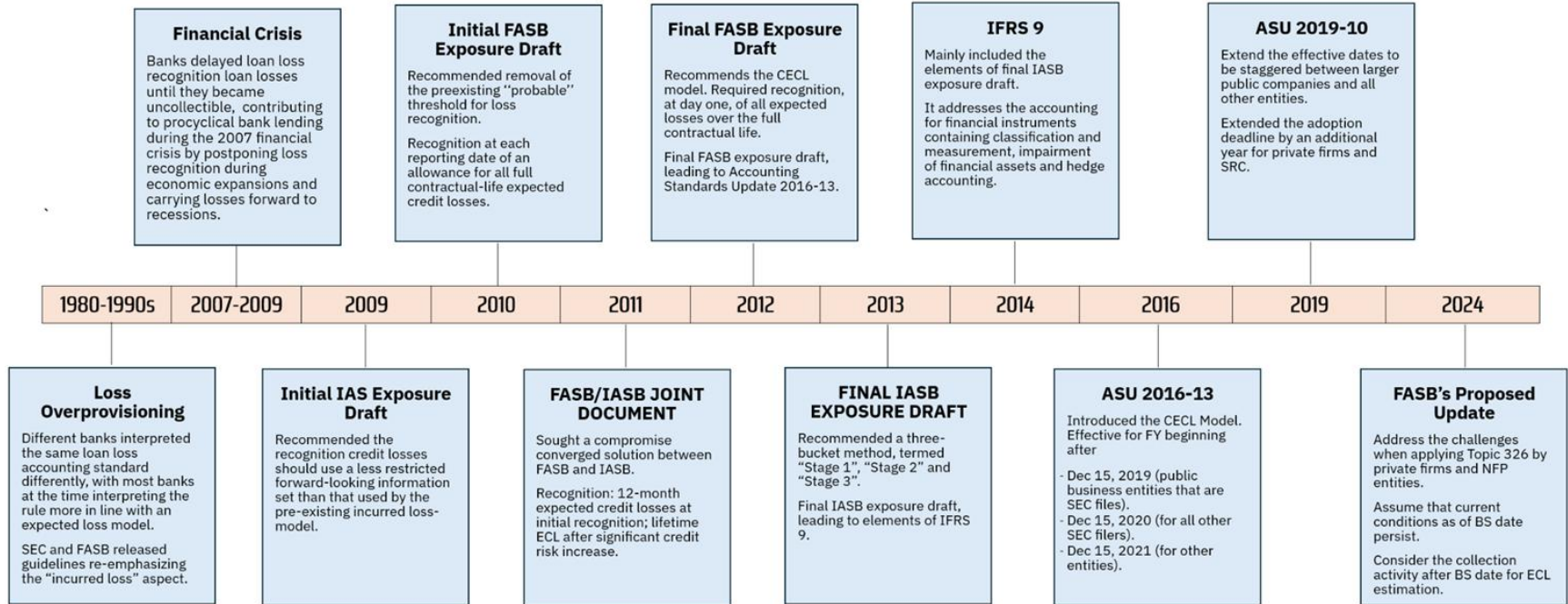
Stevens Institute of Technology

Sponsored by:

CRAFT (Center for Research toward Advancing Financial Technologies),
an **NSF-funded** Industry–University Cooperative Research Center (**I/UCRC**)

BNY Mellon | Park Avenue Finance | Wells Fargo

Background: Current Expected Credit Loss (CECL)



Incurred Loss Model vs. CECL Model

Aspect	Incurred Loss Model	CECL Model
Governing Standard	FASB's ASC 450-20 and ASC 310-10	FASB's ASC 326
Recognition Timing	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).
Data Used	Historical data and current observable conditions.	Incorporate reasonable and supportable forecasts of future economic conditions.
Forecasting Method	Simple historical loss rate analysis.	Complex, includes multiple economic scenarios.
Measurement Period	Short-term, incurred losses only.	Long-term, Entire life-of-loan (full contractual term).
Time Horizon	Short-term (12-month loss emergence).	Focuses on lifetime expected losses over the loan’s term.
Financial Impact	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.
Role of Quant Analyst Team	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

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

2. Academic Approach

1. Forward-looking measures outperform historical net charge-offs in one-year prediction with a cross-sectional regression (Harris et al., 2018).
2. Bank stock prices capture lifetime losses beyond reported allowances under vintage analysis (Wheeler, 2021).
3. 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

SEC Comment Letters

Bank	Annual Report Year	Excerpt
NY Community BankCorp	2024	“Please tell us and revise future filings to discuss in greater detail the impact that repricing risk has on your borrower’s ability to repay loans, how such risk is included in your allowance for loan losses assumptions and/or qualitative factors, and any potential negative impacts of repricing on your business.”
EWSB Bancorp	2024	“Please be as specific and detailed as needed to provide an investor with a clear understanding of any material observed changes in risk and how these changes, as well as any other key drivers, impacted each component of the allowance for loan losses established at period end with a focus on one- to four family real estate loans and consumer loans.”

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 Question

- **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?



CPE Questions

LLM-Driven AI Agent

Definition of LLM-Driven AI Agent

- Autonomous software entity designed for goal-directed task execution, with a large language model (LLM) as the core reasoning component (Sapkota et al. 2026).

Core Characteristics of AI Agent Design (Sapkota et al. 2026)

- Autonomy,
- Task-specificity, and
- Reactivity and adaptation.

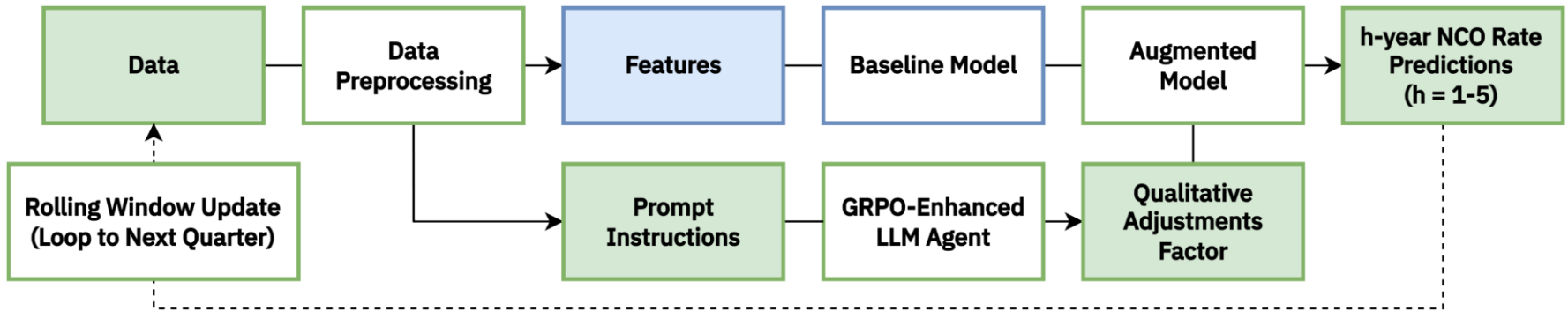
Key Implementation Challenges

- Maintain tight task-specificity,
- Ensuring qualitative adjustments respond appropriately to economic conditions (e.g., more conservative in downturns) without overreacting.



Fine-Tuning AI Agent: Group Relative Policy Optimization (GRPO) Approach

- **GRPO:** A reinforcement learning algorithm proposed by the DeepSeek team for fine-tuning large language models (Shao et al., 2024).
 - Fine-tuning is the process of updating a model's parameters using additional training data or feedback signals to improve its performance on a specific task.



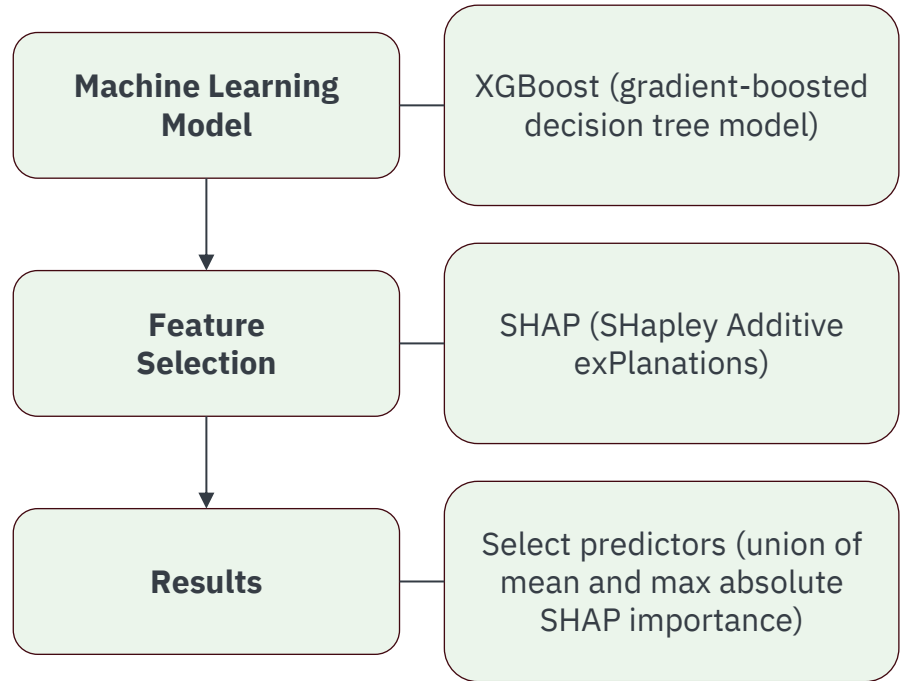
Predictors for NCO Rate

- **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.

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 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

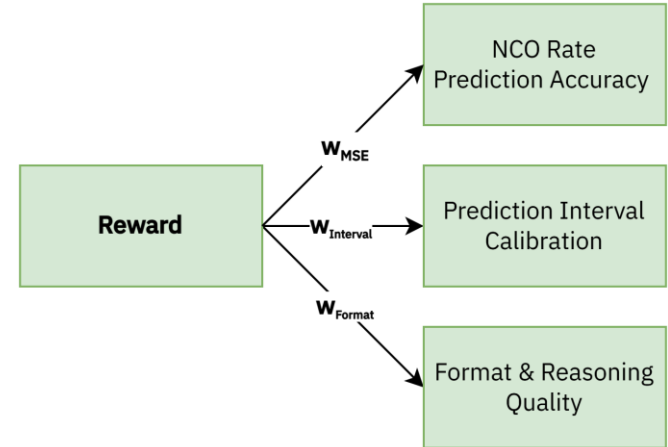
</prediction>



GRPO - Fine-Tuning Iteration

Phase 1: GRPO Fine-Tuning Data Preparation (per Round)

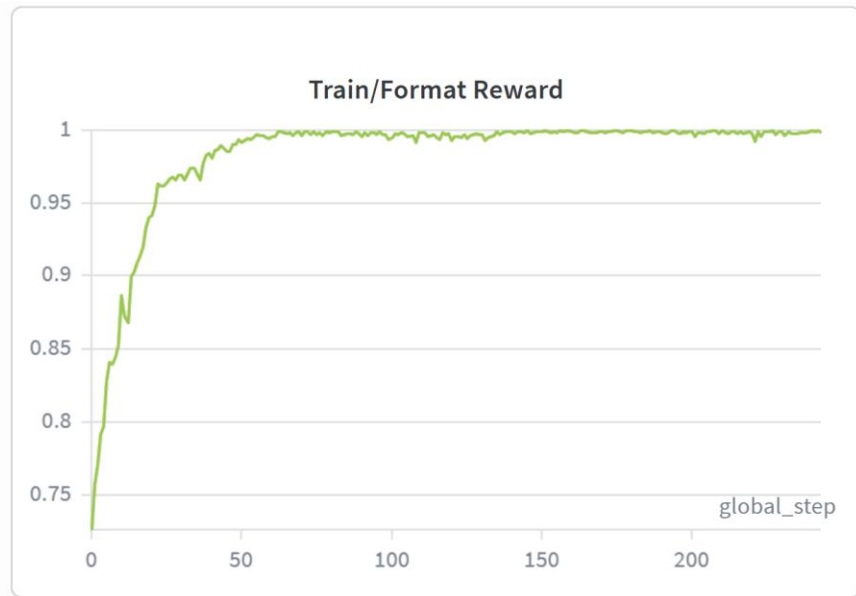
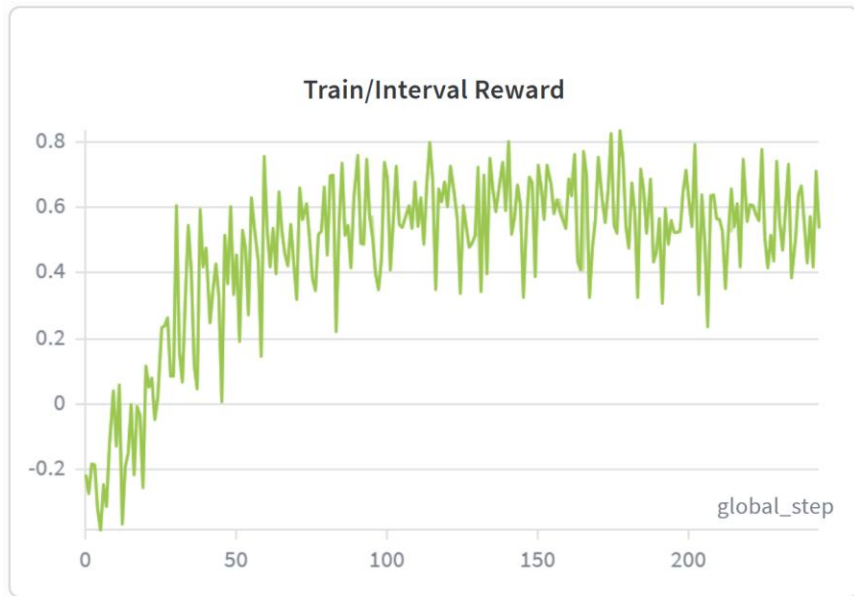
1. Provide the model with the prompt containing bank-specific content.
2. Generate multiple responses (e.g., 8) with slight randomness.
3. Evaluate each response using the reward function.
4. Compare responses with the group and determine relative quality (standardize scores).



Reward Components

Reward Component	Objective	Weight
Prediction Accuracy	Measures numeric accuracy of the next-quarter NCO prediction	0.5
Confidence Interval Calibration	Rewards well-calibrated and accurate confidence intervals of NCO rate prediction	0.3
Reasoning Quality and Formatting	Measures how thoroughly the model explains its prediction	0.2

Reward Convergence during GRPO Fine-Tuning



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.



LLM Model Results

TABLE 1: Predicting the NCO Rates of Individual Banks with LLM Models

Panel A: Out-of-Sample MSE × 1000

Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	LLM (Pre-Trained)	0.205	0.389	0.471	0.503	0.524
2	LLM (GRPO-Tuned)	0.121	0.183	0.493	0.227	0.327

Panel B: Out-of-Sample MAE × 100

Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	LLM (Pre-Trained)	0.607	0.877	0.982	1.039	1.068
2	LLM (GRPO-Tuned)	0.452	0.596	1.005	0.819	1.237

- GRPO-tuned LLM consistently outperforms the pre-trained LLM for one- and two-year-ahead predictions.
- GRPO-tuned LLM have better or comparable performance for the long-term predictions.

Baseline vs. LLM Results

TABLE 2: Predicting the NCO Rates of Individual Banks with Selected Baseline and LLM Models

Panel A: Out-of-Sample MSE $\times 1000$

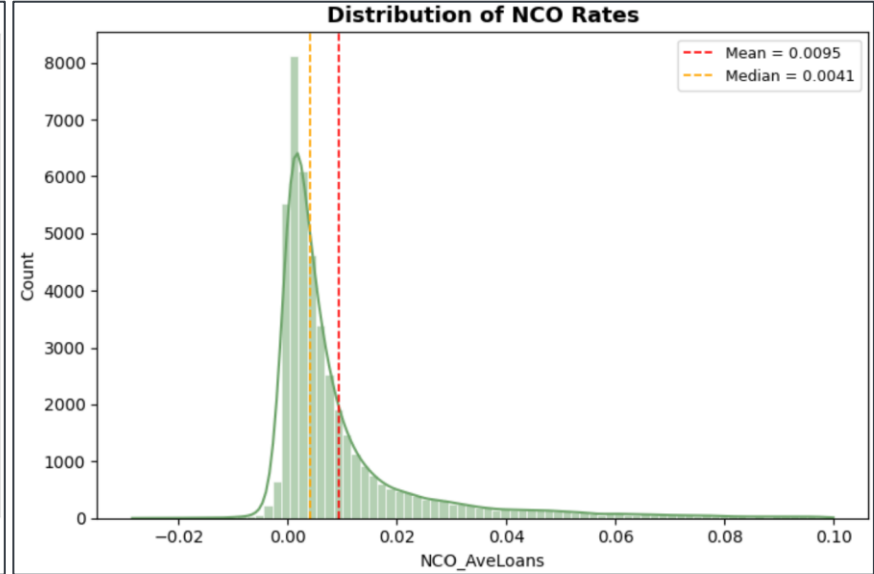
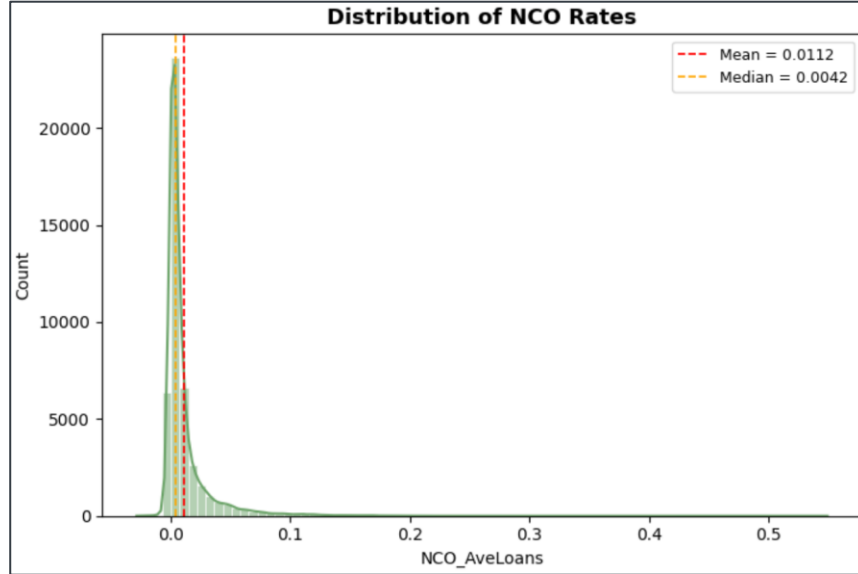
Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	HKN	0.086	0.147	0.318	1.135	1.016
2	Expanded HKN	0.229	0.221	0.223	0.242	0.959
3	LLM (Pre-Trained)	0.205	0.389	0.471	0.503	0.524
4	LLM (GRPO-Tuned)	0.121	0.183	0.493	0.227	0.327

Panel B: Out-of-Sample MAE $\times 100$

Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	HKN	0.450	0.634	0.870	1.301	1.435
2	Expanded HKN	1.187	1.163	1.211	1.293	2.096
3	LLM (Pre-Trained)	0.607	0.877	0.982	1.039	1.068
4	LLM (GRPO-Tuned)	0.452	0.596	1.005	0.819	1.237

- HKN dominates at short horizons but deteriorates at longer ones.
- GRPO-Tuned LLM offers the most balanced performance across all horizons.

Distribution of NCO Rates



Baseline vs. Augmented Model Results

TABLE 3: Predicting the NCO Rates of Individual Banks with Baseline and Augmented Models (OLS)

Panel A: MSE Improvement (%) over HKN

Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	Aug HKN (Pre-Trained) – Numeric Factor	-0.38	-0.05	0.44	0.10	-0.05
2	Aug HKN (Pre-Trained) – Directional Factor	0.21	0.10	0.70	0.46	0.47
3	Aug HKN (GRPO-Tuned) – Numeric Factor	1.18	-0.44	0.87	0.37	-0.58
4	Aug HKN (GRPO-Tuned) – Directional Factor	1.65	1.45	0.56	1.08	1.61

Panel B: MAE Improvement (%) over HKN

Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	Aug HKN (Pre-Trained) – Numeric Factor	-0.69	-0.12	0.07	-0.00	-0.03
2	Aug HKN (Pre-Trained) – Directional Factor	1.22	1.08	1.85	1.12	0.64
3	Aug HKN (GRPO-Tuned) – Numeric Factor	1.68	0.05	0.88	-0.00	-0.73
4	Aug HKN (GRPO-Tuned) – Directional Factor	3.12	2.44	0.88	0.90	0.96

Baseline vs. Augmented Model Results

TABLE 3: Predicting the NCO Rates of Individual Banks with Baseline and Augmented Models (OLS)

Panel C: MSE Improvement (%) over ExpHKN

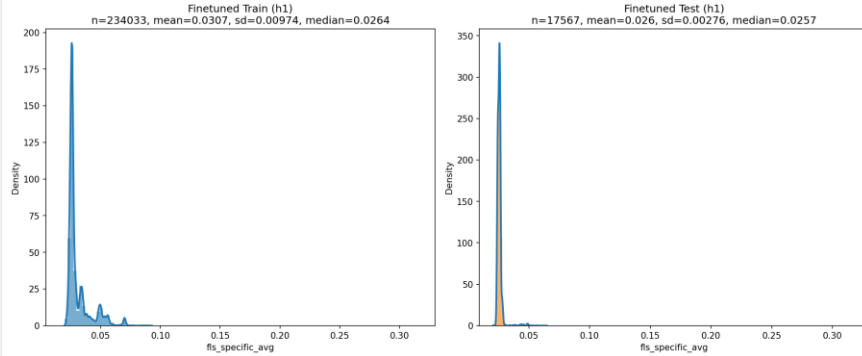
Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	Aug ExpHKN (Pre-Trained) – Numeric Factor	20.43	0.75	2.79	3.24	5.00
2	Aug ExpHKN (Pre-Trained) – Directional Factor	33.66	19.13	10.57	5.99	3.44
3	Aug ExpHKN (GRPO-Tuned) – Numeric Factor	47.56	32.28	-45.02	21.05	33.01
4	Aug ExpHKN (GRPO-Tuned) – Directional Factor	41.58	32.82	9.21	28.05	6.52

Panel D: MAE Improvement (%) over ExpHKN

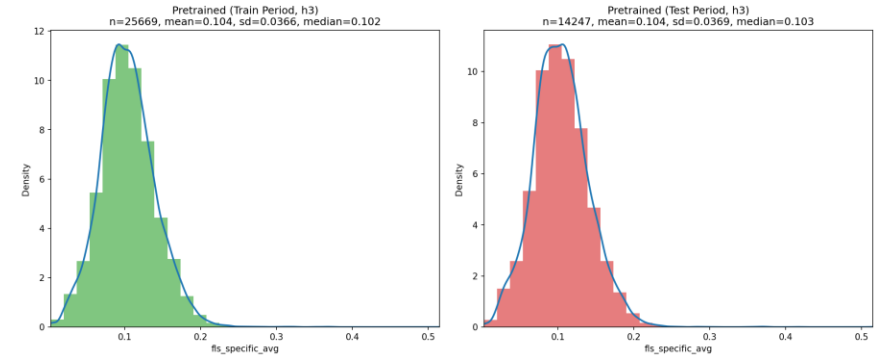
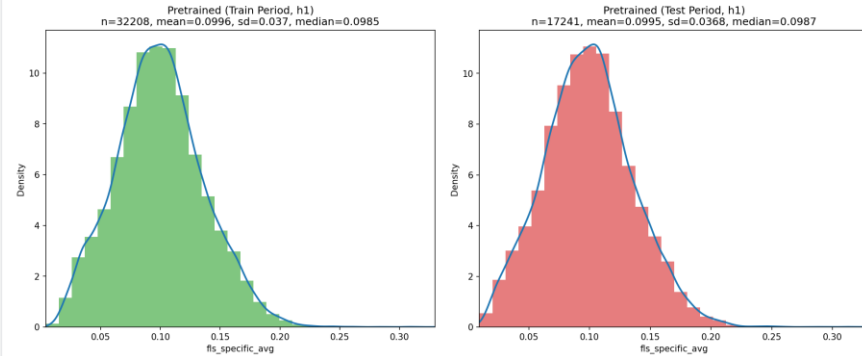
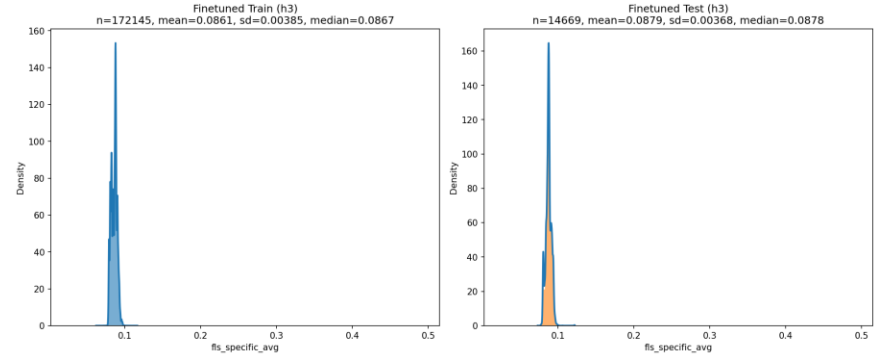
Row	Model Type	$p_{T+1 T}$	$p_{T+2 T}$	$p_{T+3 T}$	$p_{T+4 T}$	$p_{T+5 T}$
1	Aug ExpHKN (Pre-Trained) – Numeric Factor	32.27	5.36	2.48	1.79	2.63
2	Aug ExpHKN (Pre-Trained) – Directional Factor	44.45	21.67	9.07	4.45	3.86
3	Aug ExpHKN (GRPO-Tuned) – Numeric Factor	58.55	42.83	10.72	25.70	15.50
4	Aug ExpHKN (GRPO-Tuned) – Directional Factor	56.13	48.55	20.56	32.83	6.38

Distribution of FLS Specificity Scores

Distribution of Bank-Quarter FinBERT FLS-Specificity Scores (h1)



Distribution of Bank-Quarter FinBERT FLS-Specificity Scores (h3)



Takeaways

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.
- Long-term predictions and banks with unconventional business models remain a challenge.

Next Steps

- Construct augmented models by aggregating multiple LLM forecast views through a consensus-based approach.
- Improve GRPO performance through reward weight and hyperparameter optimization.
- Demonstrate how AI-generated adjustments address bounded rationality constraints.



Conclusion

- **Operationalizes CECL qualitative adjustments** through agentic reasoning—enabling systematic, auditable compliance with CECL requirements
- **Demonstrates AI agent augmentation** of traditional econometric approaches in accounting estimation

CPE Questions

Research Initiative: CECL Implementation

- Our team is examining the application of the (CECL) model from the perspectives of banks, regulators, and auditors.
- We are seeking insights and **discussions** with professionals experienced in CECL implementation and oversight.
- Please free to contact
 - Arion Cheong (acheong@stevens.edu)
 - Shanshan Yang (syang56@stevens.edu)
 - Steve Yang (syang14@stevens.edu)

We greatly appreciate your input!

Any
Question





AI and Structured Data Forum: *Optimizing Performance*

Friday, May 15 | Stevens Institute of Technology



- **AI, Audit & Assurance**

- Brian Crowley, Audit & Assurance Managing Director, Deloitte & Touche LLP
- Brian Hecker, Audit Partner, Crowe LLP
- Representative from PwC
- Moderated by Ami Beers, CPA, CGMA, Senior Director, Assurance and Advisory Innovation, AICPA

- **Interoperability in DeFi**

- Dr. Zachary Feinstein, Associate Professor in the School of Business at Stevens Institute of Technology
- Rob Krugman, Chief Digital Officer, Broadridge Financial Solutions
- Jason Meyers, Lead Architect, Creator of Pacioli.ai, Auditchain

- **Keynote:** Julie Marlowe, Office of Structured Disclosure, Division of Economic and Risk Analysis, U.S. Securities and Exchange Commission