

# MOODY'S

**Judge Moody's: Automating  
Semantic Search Relevance  
Evaluation with LLM Judges**

**Haystack 2025**

Gurion Marks

Senior Software Engineer, Search

# Roadmap

MOODY'S

1

## **RAG Systems**

Moody's Research Assistant

2

## **Search Evaluation**

Traditional methods and challenges

3

## **Solution: LLMs!**

Automated search relevance framework

4

## **Tuning**

Getting a LLM to act as an expert

5

## **Results**

Agreement with expert judges

6

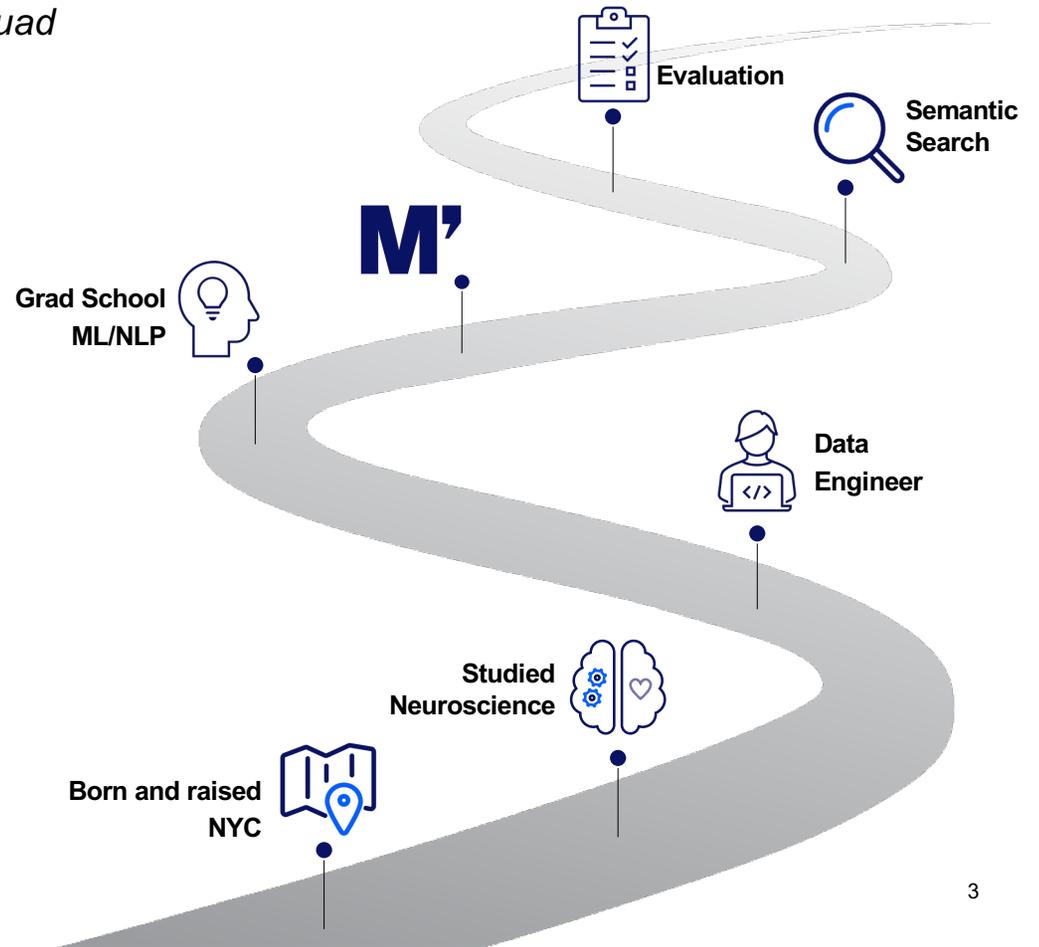
## **Ongoing Work and Lessons Learned**

Future experiments and caveats

# About Me



**Gurion Marks**  
Senior Software Engineer  
*Search Squad*





**RAG Systems:  
Moody's Research Assistant**

# What is RAG? Why is it important?

A few years ago...

## RETRO (2022)

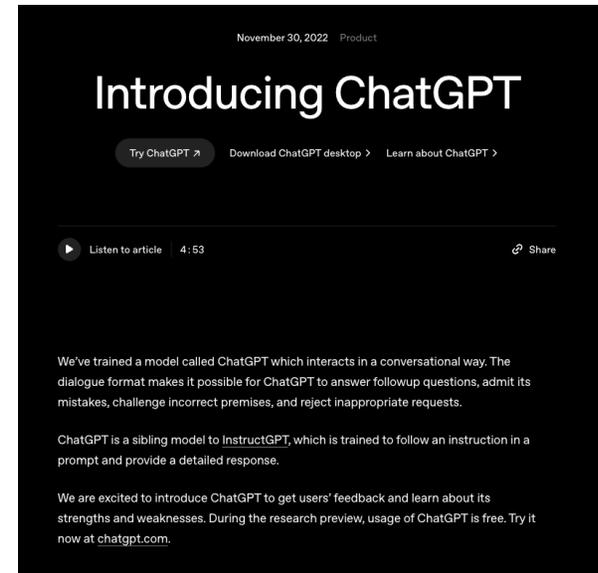


### Improving language models by retrieving from trillions of tokens

Sebastian Borgeaud<sup>1</sup>, Arthur Mensch<sup>1</sup>, Jordan Hoffmann<sup>1</sup>, Trevor Cai, Eliza Rutherford, Katie Millican, George van den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, Diego de Las Casas, Aurelia Guy, Jacob Menick, Roman Ring, Tom Hennigan, Saffron Huang, Loren Maggiore, Chris Jones, Albin Cassirer, Andy Brock, Michela Paganini, Geoffrey Irving, Oriol Vinyals, Simon Osindero, Karen Simonyan, Jack W. Rae<sup>1</sup>, Erich Elsen<sup>2</sup> and Laurent Sifre<sup>1,3</sup>  
All authors from DeepMind, <sup>1</sup>Equal contributions, <sup>2</sup>Equal senior authorship

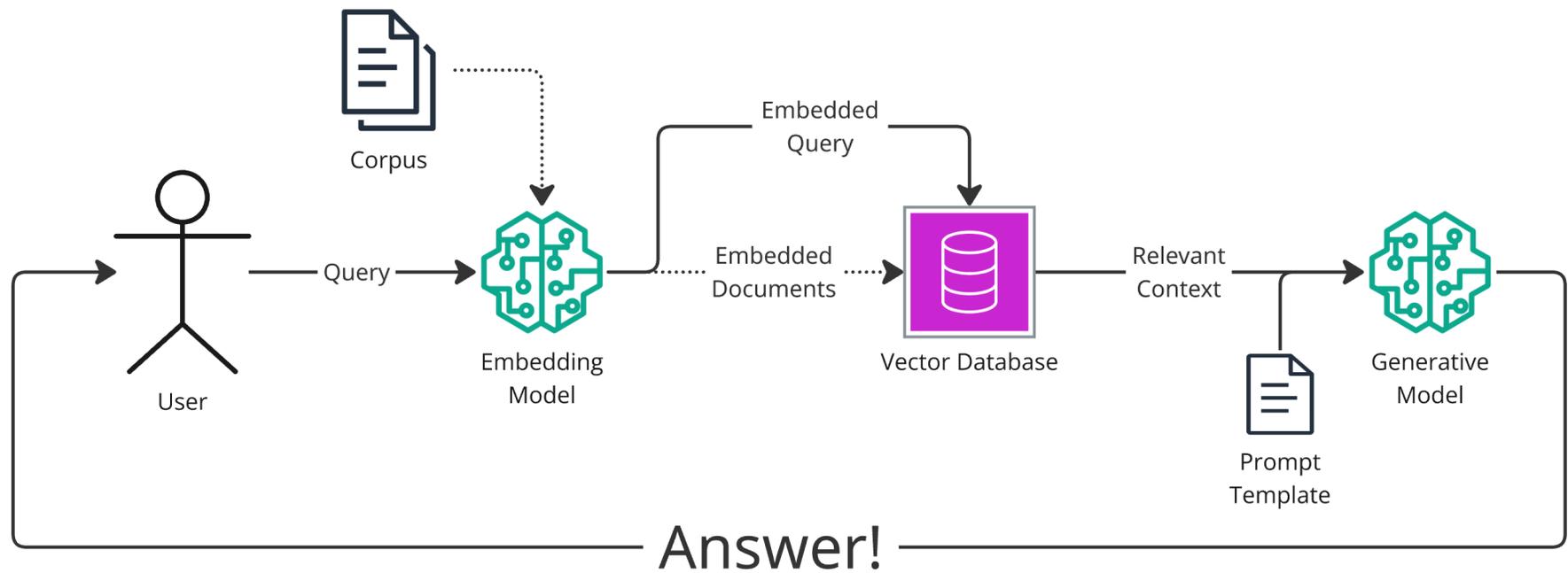
We enhance auto-regressive language models by conditioning on document chunks retrieved from a large corpus, based on local similarity with preceding tokens. With a 2 trillion token database, our Retrieval-Enhanced Transformer (RETRO) obtains comparable performance to GPT-3 and Jurassic-1 on the Pile, despite using 25x fewer parameters. After fine-tuning, RETRO performance translates to downstream knowledge-intensive tasks such as question answering. RETRO combines a frozen BERT retriever, a differentiable encoder and a chunked cross-attention mechanism to predict tokens based on an order of magnitude more data than what is typically consumed during training. We typically train RETRO from scratch, yet can also rapidly RETROfit pre-trained transformers with retrieval and still achieve good performance. Our work opens up new avenues for improving language models through explicit memory at unprecedented scale.

## ChatGPT



# What is RAG? Why is it important?

## Generic Architecture



# RAG at Moody's: Research Assistant

## General Process

### Data Sources and Flow

---

MANY steps and data sources combined

→ Intent detection and named entity recognition

- Allows filtering in retrieval step for higher quality results

– e.g. Query is “Write me a credit memo about <COMPANY> that includes <FEATURES>, top 3 peers with <FEATURES>”

→ Different searches for context

- Research
- Organizations
- News
- Other Databases

→ Combine search results and summarize context into a final answer

# Requirements on Research Assistant

What do we need to trust RA?

---

**Accurate Data**



ZERO hallucinations  
Representative of current data/trends

---

**Traceable**



Answers always linked to source  
Citations

---

The background of the slide is a solid dark blue color. Overlaid on this background are several sets of thin, light blue lines that form a complex, wavy, and somewhat chaotic pattern. These lines resemble a stylized representation of a signal or a network, with some areas being more densely packed than others. The overall effect is a modern, technical, and abstract aesthetic.

# **Search Evaluation: Traditional Methods and Challenges**

# How do we evaluate Research Assistant?

## What are we trying to evaluate?

### Final Generation

---

We want to evaluate the answers given to the user

- Can use frameworks like RAGAS to calculate metrics such as
  - Faithfulness
    - Extent to which claims are supported by retrieved context
  - Response Relevancy
    - How relevant a response is to a user input
- Uses retrieved context, but is evaluating *final results*, not *model inputs*

### Inputs to Model (Context)

---

We want to know that the generation step is even getting the correct information

- If the inputs to the generation step are lacking, we can't expect faithful results
- Basic, tried-and-true information retrieval metrics
  - Precision, Recall, etc.
- **We need a way to evaluate model context at scale**
  - **(search team focus)**

# How Do We Evaluate Research Assistant?

## How do we evaluate context?

**Problem:**

I'm a software engineer, not a finance domain expert

---

**Subject Matter Experts** → People who can say with confidence whether a chunk is relevant to a query

---

**Traditional Metrics** → Precision, recall, etc.  
Easy to compute, but we need SME review first

---

# How Do We Evaluate Research Assistant?

## Human Evaluation – Expert Evaluators

### Expert Evaluators

---

We are dealing with a complex domain in which an untrained evaluator is likely to be incorrect

→ We need finance domain experts to review our results to determine the quality of our results

- Time intensive
  - Evaluators will take days to weeks to review data
- Costly
  - Experts are highly compensated

# Metrics

## Overview of Common Metrics

### 1 Order-unaware

---

- Accuracy
  - Overall proportion of correct predictions
- Precision
  - Proportion of true positives among all predict
- Recall
  - Fraction of retrieved instances out of all relevant ones

### 2 Order-aware

---

- nDCG
  - How well a ranking system puts relevant items at the top
- mAP
  - Average precision of a set of queries



**Solution, LLMs:  
Automated Search Relevance Framework**

# Judge Moody's

## LLMs as Evaluators

### Concept

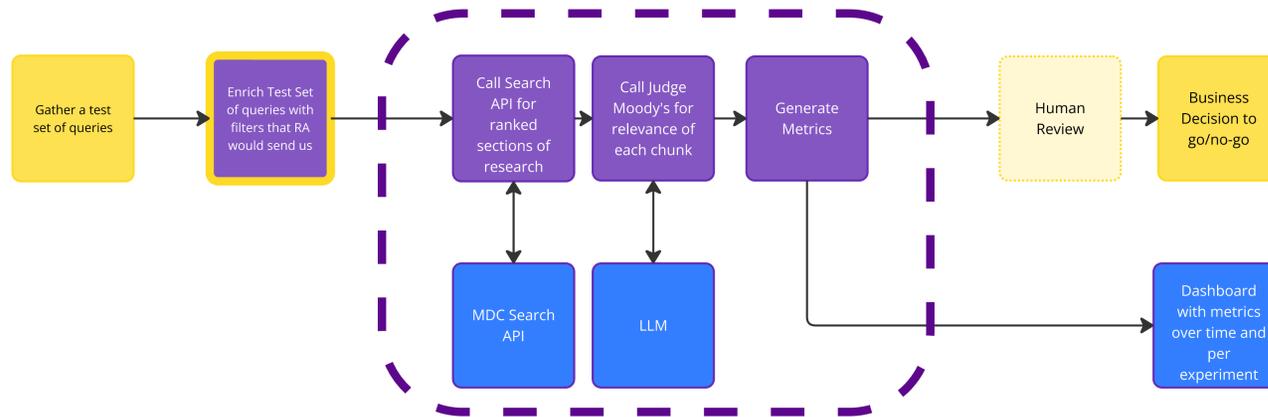
---

Perhaps a model trained on the *entire internet* could act as a reasonable relevance judge...

- They probably learned some finance during their pretraining
- With enough examples they could understand the task
- Language models are much *faster* than humans
- Language models are much *cheaper* than humans

# Implementing Judge Moody's

## Architecture



## Lower-level details

- Sample prompts stored in S3 to make iteration/templating easy
- Precomputed judgments stored to prevent re-running the same query-chunk pairs
  - Important cost savings when smaller algorithm changes are made and many of the same chunks are returned
- Results discussed in this presentation use GPT-4o, but as new foundation models are released we test those as well

The background of the slide is a deep blue color. It features a complex, abstract pattern of thin, light blue lines that form a series of overlapping, wavy, and mesh-like structures. These lines create a sense of depth and movement, resembling a digital or neural network aesthetic. The pattern is most prominent on the left side and fades slightly towards the right.

# **Tuning: Getting a LLM to Act as an Expert**



**Please tell me  
whether the following  
chunk of text is  
relevant to answering  
this query:**

**Query: <query>**

**Chunk: <chunk>**

# Prompt Engineering Techniques

## Role Prompting

### Tell the model about itself

---

Highly tuned edits on the model's "persona" and "role"

- Tell the model it's a domain expert *AS WELL AS* a software engineer/search expert
- Give the LLM a "setting" in which it's working
  - "an intern did a task for you that you need to review"



### **# Role**

**You are a finance domain expert working at Moody's Ratings and consult with the search team to make sure your research is readily accessible.**

### **# Task**

**Your intern built a semantic search algorithm to return text chunks from financial research documents, and you need to evaluate this algorithm to see how relevant the retrieved chunks are to answering a query.**

**Review this example pair:**

**Query: <query>**

**Chunk: <chunk>**

# Prompt Engineering Techniques

## Explicit Evaluation Criteria

### Give a rubric to follow

---

The more specific your instructions, the more the model will be able to follow them

- Give multiple aspects, or *subtasks*, with weights that need to be attended to
  - Within the larger overall task, weight 70% of your evaluation on feature A, and 30% on feature B
- Give a pointing system for each subtask
  - Give definitions for what good results may contain
  - Make sure to address and discuss edge cases that would trip you up as a human
  - The clearer and more explicit your scoring criteria, the better performance on your task



### **# Role**

**You are a finance domain expert working at Moody's Ratings and consult with the search team to make sure your research is readily accessible.**

### **# Task**

**Your intern built a semantic search algorithm to return text chunks from financial research documents, and you need to evaluate this algorithm to see how relevant the retrieved chunks are to answering a query.**

### **# Scoring Criteria**

#### **## Rubric**

- 0: The chunk is irrelevant to the query or you are unable to assess relevance based only on the chunk**
- 1: The document is relevant to the query and provides information useful to answering the question**

#### **## Considerations**

- If you do not understand a term do not try to guess**
- Be strict in your assessment and if you are unsure mark as irrelevant**

**Review this example pair:**

**Query: <query>**

**Chunk: <chunk>**

# Prompt Engineering Techniques

## Few-Shot Learning

### Give examples of what results should look like

---

*Language Models are Few-Shot Learners* (Brown et al, 2020)

- Giving examples improves quality
  - Positive and negative cases
  - Edge cases and what results should look like
- Make the model explain its reasoning prior to answering the prompt
  - Make the model “think” and have reasoning in its context window for producing the final answer
- Give an output format to ensure results conform to what you’re expecting
  - Structured JSON, precise values for results, etc.
  
- *As reasoning models have gotten better, these seem to be having less impact*



**# Role**  
**# Task**  
**# Scoring Criteria**

**# Examples**

**## Input**

**Query: <query>**

**Chunk: <chunk>**

**## Output**

**{**

**explanation: The document is relevant to the query because...**

**judgment: 1**

**}**

**Review this example pair and return your explanation and judgment in JSON format:**

**Query: <query>**

**Chunk: <chunk>**

**# Output**

**Results:**

**Agreement with Expert Judges**



# Experiment Setup

## Judge Moody's vs. Domain Experts

**8** Expert Evaluators

---

Finance Domain Experts with decades of experience in the field

**75** Example Pairs

---

Real query-chunk pairs taken from Research Assistant queries and context

---

### Task

- Given user question and retrieved research passage
  - Does provided chunk help answer the question (Yes/No)
- Judge Moody's judgment shown
  - If expert judgment differs, explain reasoning

# Do We Really *NEED* Experts?

## Judge Moody's vs. The Experts #1

---

**Query** Analyze [Insurance Company]. What are the company's operational risks and how does the company mitigate its operational risks?

**Chunk** [Insurance Company]: Large Midwest personal lines insurer focused on improving weak autoresults  
Profitability: Underwriting actions and integration underway to reduce combined ratios and improve results: [Insurance Company] mutual holding company structure provides some competitive advantages over its publicly traded competitors since [Insurance Company] can pursue customer-driven initiatives while withstanding volatility of insurance cycles without the concern of meeting return on capital expectations of shareholders. The company's statutory expense ratios are generally higher than its personal lines peer group. [Insurance Company] is enhancing technology to streamline processes to improve efficiency. We expect the company to integrate its past acquisitions and lower its expense ratio to compete with well capitalized national carriers and low-cost direct response personal line insurers.

---

### Judge Moody's: FALSE

The text chunk primarily discusses [Insurance Company] profitability, competitive advantages, and efforts to improve efficiency through technology and integration. While these efforts may indirectly relate to operational risk mitigation, the text does not explicitly address operational risks or specific risk mitigation strategies. The focus is more on financial performance and competitive positioning rather than operational risk management.

### Expert 1: True

This section discuss statutory expenses and their anticipated improvement, which indirectly address company's operational risk. However this information is relevant to competitive risk rather than operational risk.

### Expert 2: True

The piece mentions (i) higher expense ratio and measures to reduce the same and (ii) technology enhancements. Both factors relate to operational challenges

### Expert 3: True

The response provided is correct , however two statements are available in latest report which are actual are available in latest report of 2024, which are shown as shown as forecast in this 2023 report.

# Do We Really *NEED* Experts?

## Judge Moody's vs. The Experts #2

---

**Query** Give me a swot analysis on [Car Company]. Give me the answer in japanese

**Chunk** **[Car Company]: Update to credit analysis**  
Margin improvement to continue with unit recovery, cost control, and a weak yen: The improved profitability in recent quarters can also be partially attributed to the effective actions taken to mitigate the effects of the pandemic, including cost reductions and a focus on core markets and models. Its restructuring efforts, combined with low inventory because of supply chain issues, have reduced the company's incentives in the US. Along with its sales efforts in the US, [Car Company] has also been reducing its production capacity globally to rightsize its operations and focus on profitability over scale. In addition to optimizing manufacturing, and reducing marketing and sales costs, the company's focus on core models and the consolidation of core regions reduced costs from pre-pandemic levels significantly.

---

### Judge Moody's: **FALSE**

The text chunk provides information on [Car Company] recent profitability improvements, cost control measures, and restructuring efforts. However, it does not directly address a SWOT analysis, which would require a more comprehensive overview of strengths, weaknesses, opportunities, and threats. The text is more focused on specific operational strategies rather than a broad strategic analysis. Therefore, it is not highly relevant to the question seeking a SWOT analysis.

### Expert 1: **True**

The mentioned points constitute a part of SWOT analysis.

### Expert 2: **False**

The response is not from latest report. The Statement could have picked from similar section in latest report

### Expert 3: **True**

has relevant SWOT details

# Do We Really *NEED* Experts?

## Judge Moody's vs. The Experts #3

---

Query **[Bank] credit profile**

Chunk **[Bank] & Company: Update to credit analysis following the ratings affirmation**  
Methodology and scorecard: About Moody's Bank Scorecard: Our Scorecard is designed to capture, express and explain in summary form our Rating Committee's judgment. When read in conjunction with our research, a fulsome presentation of our judgment is expressed. As a result, the output of our Scorecard may materially differ from that suggested by raw data alone (though it has been calibrated to avoid the frequent need for strong divergence). The Scorecard output and the individual scores are discussed in rating committees and may be adjusted up or down to reflect conditions specific to each rated entity. Overall, the Scorecard Calculated BCA range for [Bank] is a1 to a3. The assigned BCA of a2 is at the middle of the Scorecard Calculated BCA range.

---

### Judge Moody's: True

The text chunk is relevant to the question as it provides specific information about [Bank]'s credit profile, particularly its Baseline Credit Assessment (BCA) range and assigned BCA. This directly relates to the credit analysis and financial health of [Bank], which is what the question is seeking. The methodology and scorecard explanation also provide context on how the credit profile is determined, which is pertinent to understanding [Bank]'s credit standing.

### Expert 1: False

Except for the statement "the Scorecard Calculated BCA range for [Bank] is a1 to a3," the rest of the paragraph is unrelated to the query.

### Expert 2: False

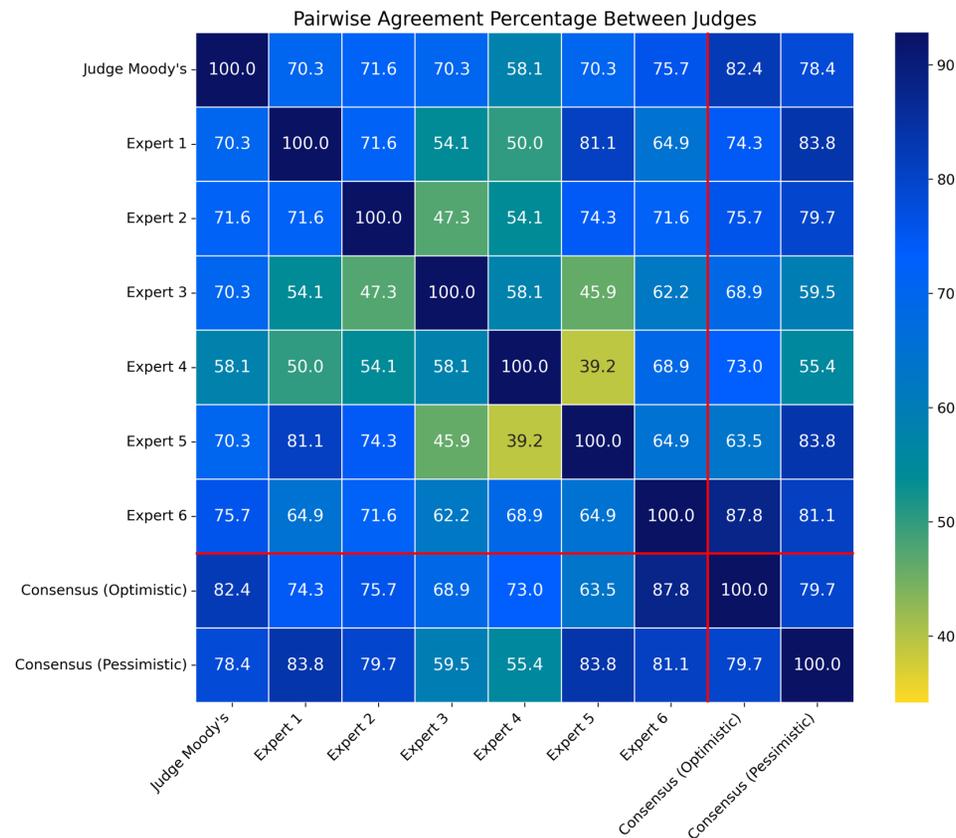
The piece talks about the methodology and no specifics about [Bank]'s credit profile

### Expert 3: False

The response provided is correct, however two statements are available in latest report which are actual are available in latest report of 2024, which are shown as shown as forecast in this 2023 report.

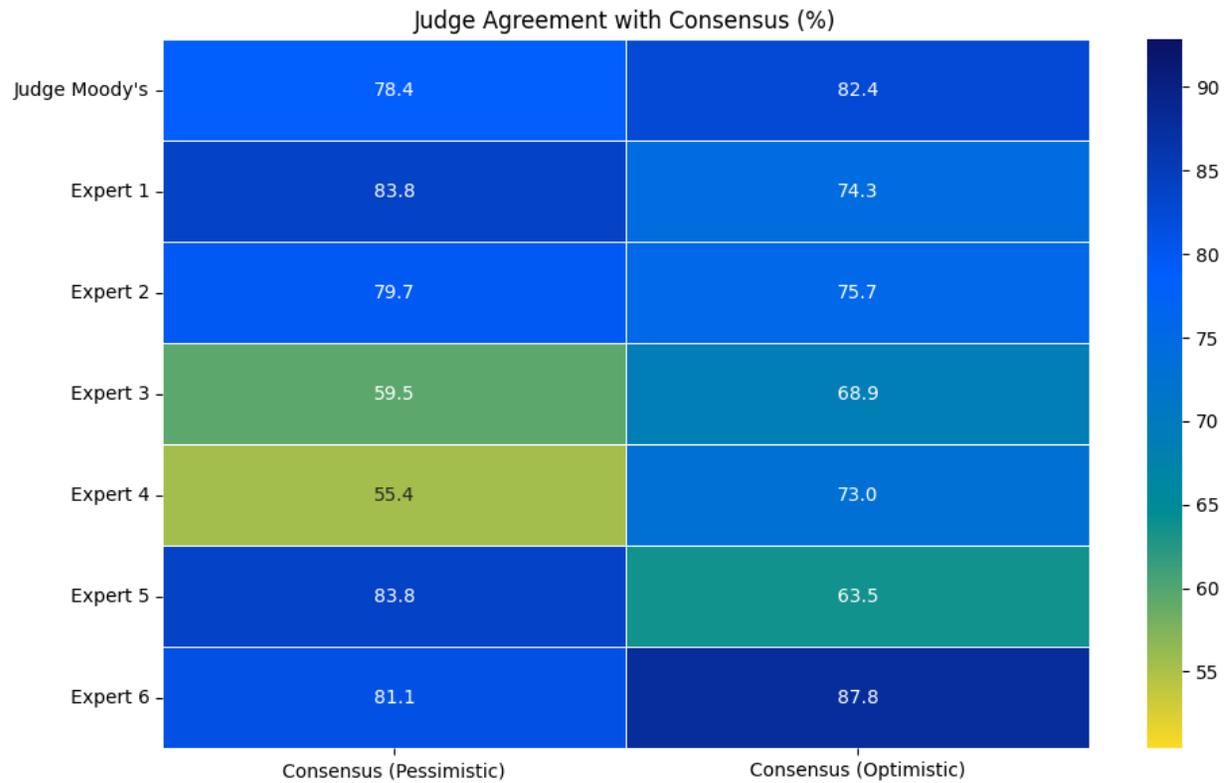
# Judge Moody's vs Expert Evaluators

## Inter-rater agreement



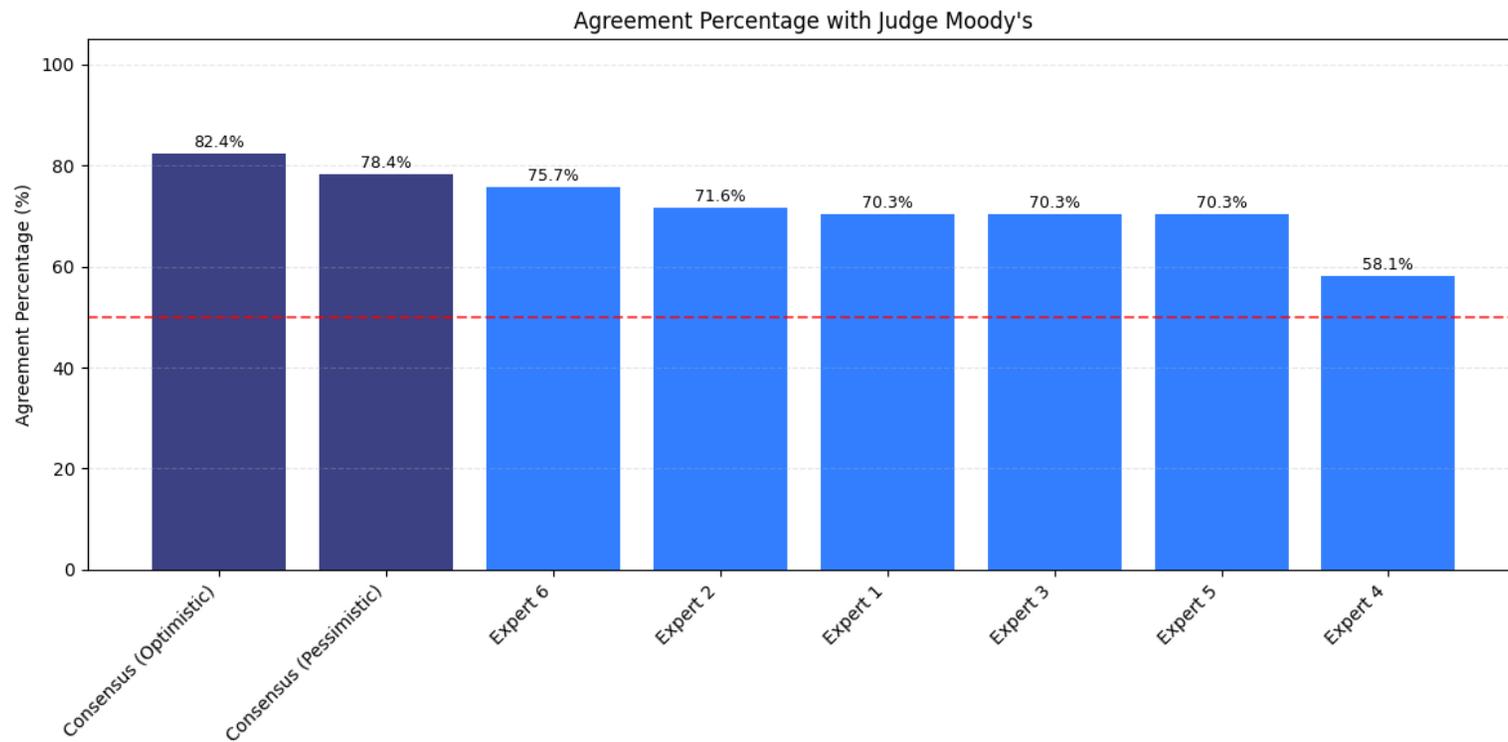
# Judge Moody's vs Expert Evaluators

## Expert Agreement with Consensus



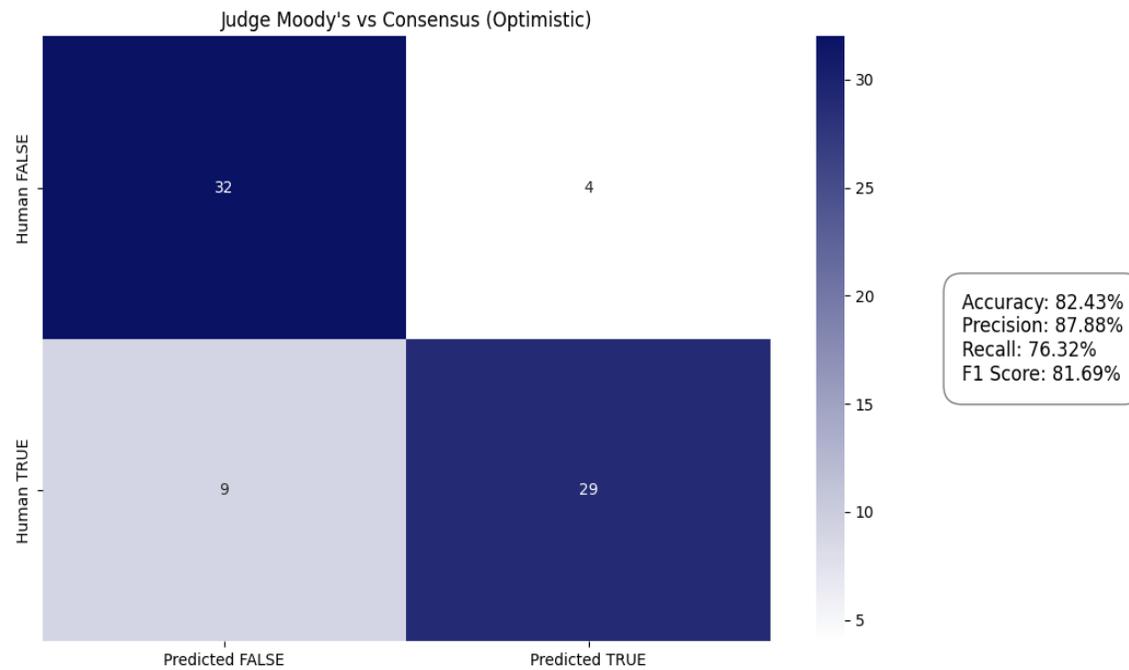
# Judge Moody's vs Expert Evaluators

## Expert Agreement with Judge Moody's



# Judge Moody's vs Expert Evaluators

## False Positives and Negatives



# Judge Moody's vs Expert Evaluators

## Time and Cost

### Time

8 Domain Experts:

- 75 query-document pairs
  - 2 week turnaround
  - 2 evaluators didn't finish all examples

Judge Moody's:

- 75 query-document pairs
  - ~3s per judgment
    - **~2 minutes**
  - 100% completion rate

### Cost

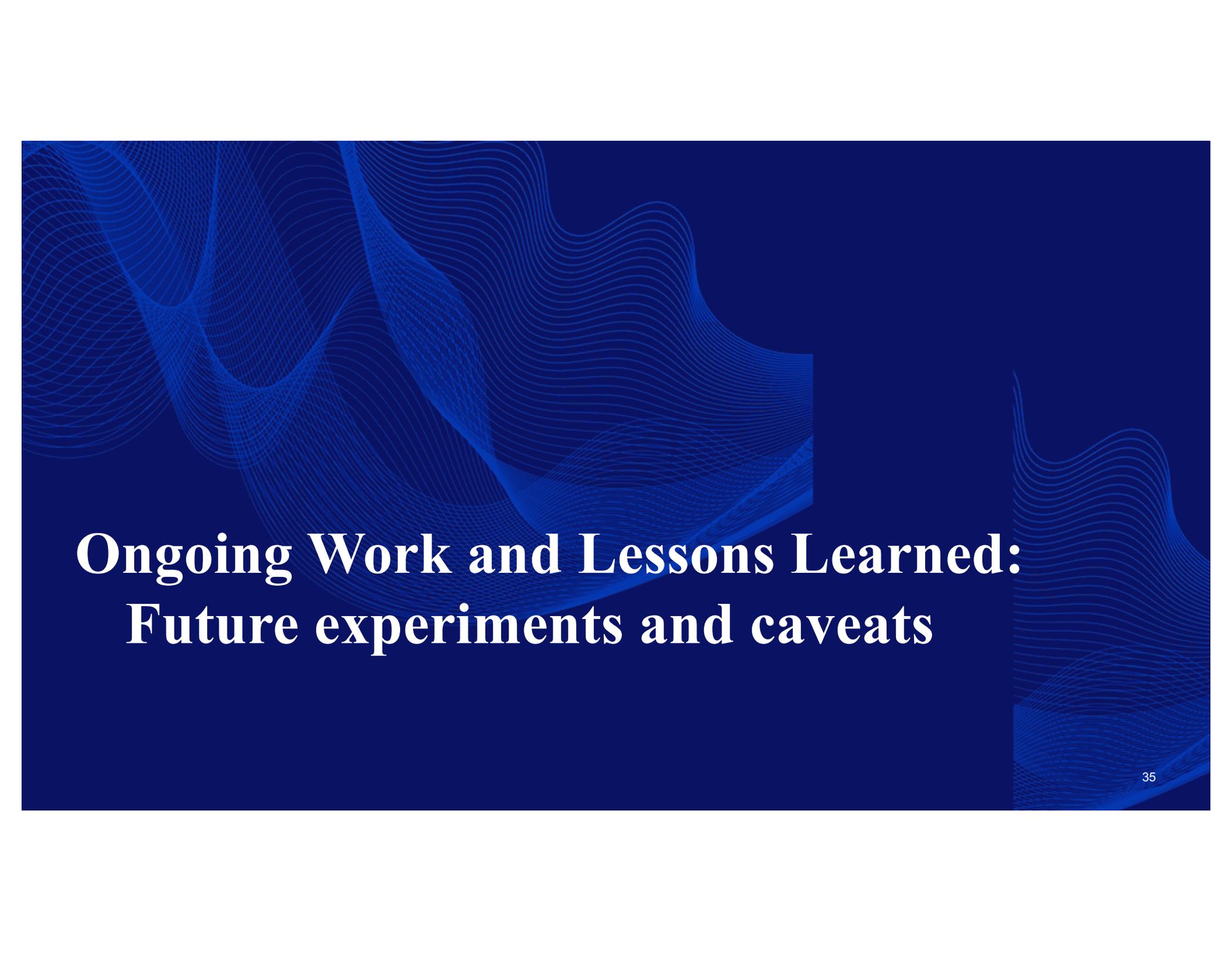
8 Domain Experts

- \$\$\$
  - Highly compensated time
  - Not doing other normal job tasks during that time

Judge Moody's

- \$
  - ~2k tokens per example, 250 token output *max* (usually closer to 100)
  - Cheap, but still may be non-trivial at scale

Model	\$/1M input	\$/1M output	Total
Claude 3.5 Haiku	\$0.80	\$4.00	\$0.18
o3-mini	\$1.10	\$4.40	\$0.23
<b>GPT-4o</b>	<b>\$2.50</b>	<b>\$10.00</b>	<b>\$0.53</b>
Claude Sonnet 3.7	\$3.00	\$15.00	\$0.68

The background of the slide is a dark blue color with abstract, wavy, light blue lines that create a sense of movement and depth. These lines are composed of many thin, overlapping curves that flow across the frame.

# **Ongoing Work and Lessons Learned: Future experiments and caveats**

# Caveats

## Trustworthiness

### Can we let this loose?

---

We don't have 100% agreement with experts

- To what extent do we trust this model to make business decisions?
  - Not even the experts agree with each other all the time
  - We can use this as a tool to iterate faster and decide when to use experts
    - We expect the model to behave predictably, so we can compare across use cases
- The biggest value-add is the ability to *iterate faster*, and reach out to domain experts when we really need them
  - We can do small-scale experiments, and determine early on which ones are worth pursuing

# More Prompt Updates!

## Taking advantage of more data

### Query Intent and Metadata

---

Update prompt to include more information gathered from user feedback and preliminary steps

→ If we know a chunk is supposed to contain information about some organization, add that as a condition to the prompt

- Query intent (topics)
- Organizations
- Etc.

# Model and Architecture Updates

## Fine-tuning and Multiple Judges

### Synthetic Data

---

Move away from LLM towards SLM:

- Train a small encoder for this NLI task
- BERT-based models are very good at logical entailment
  - Frame query-chunk relation as premise-conclusion
    - Use outputs of LLM as inputs to train a small language model for this task
    - Inspiration from *e-SNLI: Natural Language Inference with Natural Language Explanations* (Camburu, 2018)
- Potential to bring down costs for an experiment even further

### Multiple Judges

---

We see that experts don't always agree with each other but more frequently agree with consensus

- Use multiple prompts or multiple model calls (with some temperature)
- Ensemble mimics a panel of judges, rather than a single prompt that may have some bias
  - Use different prompts tuned towards different aspects
    - Perhaps a different prompt per domain
  - Ensembling raises costs (multiple of number of judges)
- Recreate behavior like what we saw in our judge panel

# Automated Prompt Tuning

## Taking advantage of expert judgments and reasoning

### Quasi-ML Approach to Prompt Engineering

---

We have a dataset of expert judgments with explanations

- Iterate on the relevance judge prompt
  - Split train-test sets
  - Run all query-chunk pairs through prompt
    - Highlight failures: use expert explanations from where model result differs from expert in a new “prompt updating prompt”
  - Set stopping condition (agreement percentage, number of iterations)
- Preliminary results show promise
  - Need more expert judgments and explanations to verify that results generalize properly

The background of the slide is a deep blue color. It features a complex, abstract pattern of thin, light blue lines that form a series of overlapping, wavy, and undulating shapes. These lines create a sense of depth and movement, resembling a digital or organic structure. The overall effect is a textured, layered appearance that changes as the viewer's perspective shifts.

# Summary

# LLM as a Judge?

What have we learned?

MOODY'S

## Large Language Model as relevance evaluator performs nearly as well as expert evaluators

- Prompt tuning is very important – small changes yield large performance impacts
- High agreement with experts, and experts don't always agree!

## Prompt Tuning Techniques

- Role Prompting
- Rubric Grading
- Few-shot Learning

## Takeaways

- Tool to iterate faster and only use experts when truly necessary
- Huge cost and time savings on a given set of relevance judgments
- Foundation models work well, but there is room for improvement potentially using fine-tuned domain-specific models

**Sandeep  
Gore**  
Tech Lead,  
Search

**Jeff  
Capobianco**  
Product Manager,  
Search

**Shuping  
Zhang**  
Director,  
Search

**David  
Fisher**  
OSC

**Shout Outs**  
*It takes a village...*

**MOODY'S**

**Thank you!**

**MOODY'S**

Questions?

MOODY'S

© 2025 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S CREDIT RATINGS AFFILIATES ARE THEIR CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND MATERIALS, PRODUCTS, SERVICES AND INFORMATION PUBLISHED OR OTHERWISE MADE AVAILABLE BY MOODY'S (COLLECTIVELY, "MATERIALS") MAY INCLUDE SUCH CURRENT OPINIONS. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT OR IMPAIRMENT. SEE APPLICABLE MOODY'S RATING SYMBOLS AND DEFINITIONS PUBLICATION FOR INFORMATION ON THE TYPES OF CONTRACTUAL FINANCIAL OBLIGATIONS ADDRESSED BY MOODY'S CREDIT RATINGS. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS, NON-CREDIT ASSESSMENTS ("ASSESSMENTS"), AND OTHER OPINIONS INCLUDED IN MOODY'S MATERIALS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S MATERIALS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. AND/OR ITS AFFILIATES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS DO NOT COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS, ASSESSMENTS AND OTHER OPINIONS AND PUBLISHES OR OTHERWISE MAKES AVAILABLE ITS MATERIALS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS, AND MATERIALS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS AND INAPPROPRIATE FOR RETAIL INVESTORS TO USE MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS OR MATERIALS WHEN MAKING AN INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT. FOR CLARITY, NO INFORMATION CONTAINED HEREIN MAY BE USED TO DEVELOP, IMPROVE, TRAIN OR RETRAIN ANY SOFTWARE PROGRAM OR DATABASE, INCLUDING, BUT NOT LIMITED TO, FOR ANY ARTIFICIAL INTELLIGENCE, MACHINE LEARNING OR NATURAL LANGUAGE PROCESSING SOFTWARE, ALGORITHM, METHODOLOGY AND/OR MODEL.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS ARE NOT INTENDED FOR USE BY ANY PERSON AS A BENCHMARK AS THAT TERM IS DEFINED FOR REGULATORY PURPOSES AND MUST NOT BE USED IN ANY WAY THAT COULD RESULT IN THEM BEING CONSIDERED A BENCHMARK.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the credit rating process or in preparing its Materials.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of

such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY CREDIT RATING, ASSESSMENT, OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

Moody's Investors Service, Inc., a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by Moody's Investors Service, Inc. have, prior to assignment of any credit rating, agreed to pay Moody's Investors Service, Inc. for credit ratings opinions and services rendered by it. MCO and all MCO entities that issue ratings under the "Moody's Ratings" brand name ("Moody's Ratings"), also maintain policies and procedures to address the independence of Moody's Ratings' credit ratings and credit rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold credit ratings from Moody's Investors Service, Inc. and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at [ir.moody's.com](http://ir.moody's.com) under the heading "Investor Relations — Corporate Governance — Charter and Governance Documents - Director and Shareholder Affiliation Policy."

Moody's SF Japan K.K., Moody's Local AR Agente de Calificación de Riesgo S.A., Moody's Local BR Agência de Classificação de Risco LTDA, Moody's Local MX S.A. de C.V., I.C.V., Moody's Local PE Clasificadora de Riesgo S.A., and Moody's Local PA Clasificadora de Riesgo S.A. (collectively, the "Moody's Non-NRSRO CRAs") are all indirectly wholly-owned credit rating agency subsidiaries of MCO. None of the Moody's Non-NRSRO CRAs is a Nationally Recognized Statistical Rating Organization.

Additional terms for Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail investors.

Additional terms for India only: Moody's credit ratings, Assessments, other opinions and Materials are not intended to be and shall not be relied upon or used by any users located in India in relation to securities listed or proposed to be listed on Indian stock exchanges.

Additional terms with respect to Second Party Opinions and Net Zero Assessments (as defined in Moody's Ratings Rating Symbols and Definitions): Please note that neither a Second Party Opinion ("SPO") nor a Net Zero Assessment ("NZA") is a "credit rating". The issuance of SPOs and NZAs is not a regulated activity in many jurisdictions, including Singapore. JAPAN: In Japan, development and provision of SPOs and NZAs fall under the category of "Ancillary Businesses", not "Credit Rating Business", and are not subject to the regulations applicable to "Credit Rating Business" under the Financial Instruments and Exchange Act of Japan and its relevant regulation. PRC: Any SPO: (1) does not constitute a PRC Green Bond Assessment as defined under any relevant PRC laws or regulations; (2) cannot be included in any registration statement, offering circular, prospectus or any other documents submitted to the PRC regulatory authorities or otherwise used to satisfy any PRC regulatory disclosure requirement; and (3) cannot be used within the PRC for any regulatory purpose or for any other purpose which is not permitted under relevant PRC laws or regulations. For the purposes of this disclaimer, "PRC" refers to the mainland of the People's Republic of China, excluding Hong Kong, Macau and Taiwan.