How might we...

ChatWatch

How might we develop ethical AI tools, like ChatGPT, to help business clients while ensuring data privacy?

ChatWatch

Who: Highly regulated employers of white collar employees in the financial services industry.

Why: Empower employees to communicate safely and in line with company values.

What: Scanner to detect and suggest fixes for compliance violations in enterprise communication channels.

ChatWatch in Action

Experiment 1: Satisfying General Council

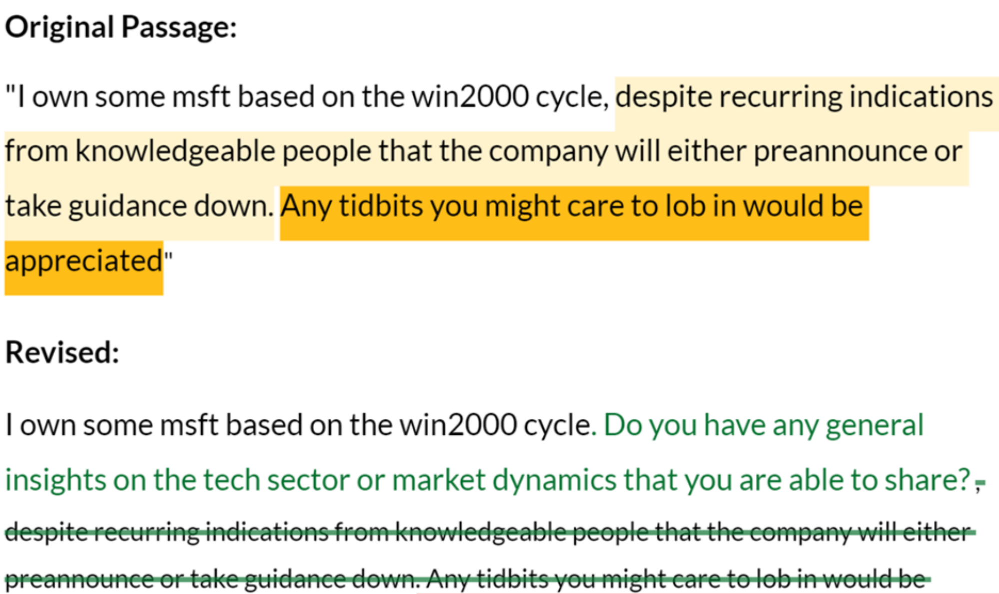
Goal: Measure the ability of the model to **catch legal issues** as well as law students
Who: JD and LLM Students from Cornell
Success Metric: % who rate model annotations as **comparable/better vs human annotation**

≟

CORNELL UNIVERSITY FOUNDED A.D. 1865

Cornell Law School

Experiment 1: Editor Perspective

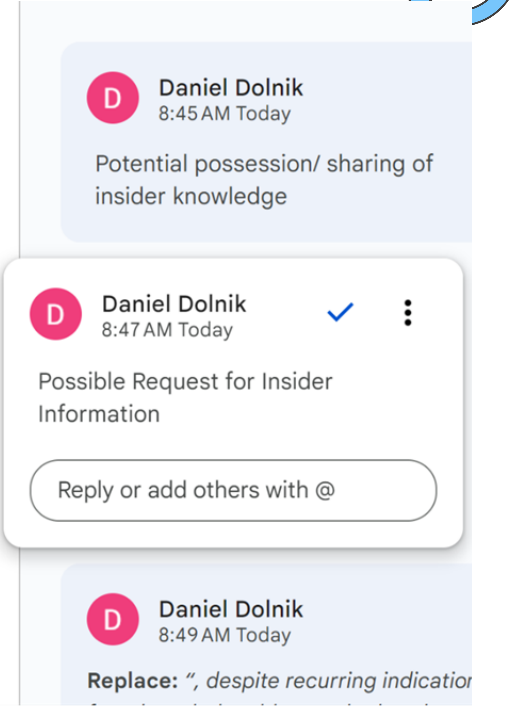


Original Passage:

"I own some msft based on the win2000 cycle, despite recurring indications from knowledgeable people that the company will either preannounce or take guidance down. Any tidbits you might care to lob in would be appreciated"

Revised:

I own some msft based on the win2000 cycle. Do you have any general insights on the tech sector or market dynamics that you are able to share?~~,~~ ~~despite recurring indications from knowledgeable people that the company will either preannounce or take guidance down. Any tidbits you might care to lob in would be appreciated.~~



Daniel Dolnik
8:45 AM Today

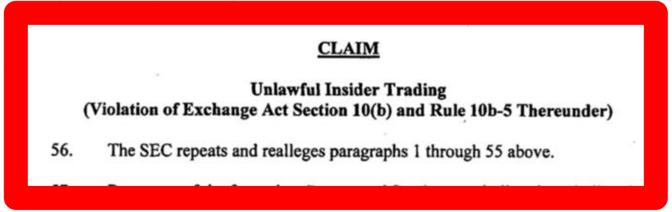
Potential possession/ sharing of insider knowledge

Daniel Dolnik
8:47 AM Today

Possible Request for Insider Information

Daniel Dolnik
8:49 AM Today

Replace: *", despite recurring indicatio*



CLAIM

Unlawful Insider Trading
(Violation of Exchange Act Section 10(b) and Rule 10b-5 Thereunder)

56. The SEC repeats and realleges paragraphs 1 through 55 above.

Experiment 2: Minimum Tolerable Performance

Goal: Measure the tolerance of the users to different levels of perceived performance
Who: Prior work experience in financial services with onboarding training
Success Metric: % of users that toggle off ChatWatch at each FP/ FN rate

Experiment 2: False positives

✕

[Level 1 False Positive]
Reason : This might be grossly exaggerating the situation consider using a slightly subdued tone
Suggestion : relatively higher than normal

✓

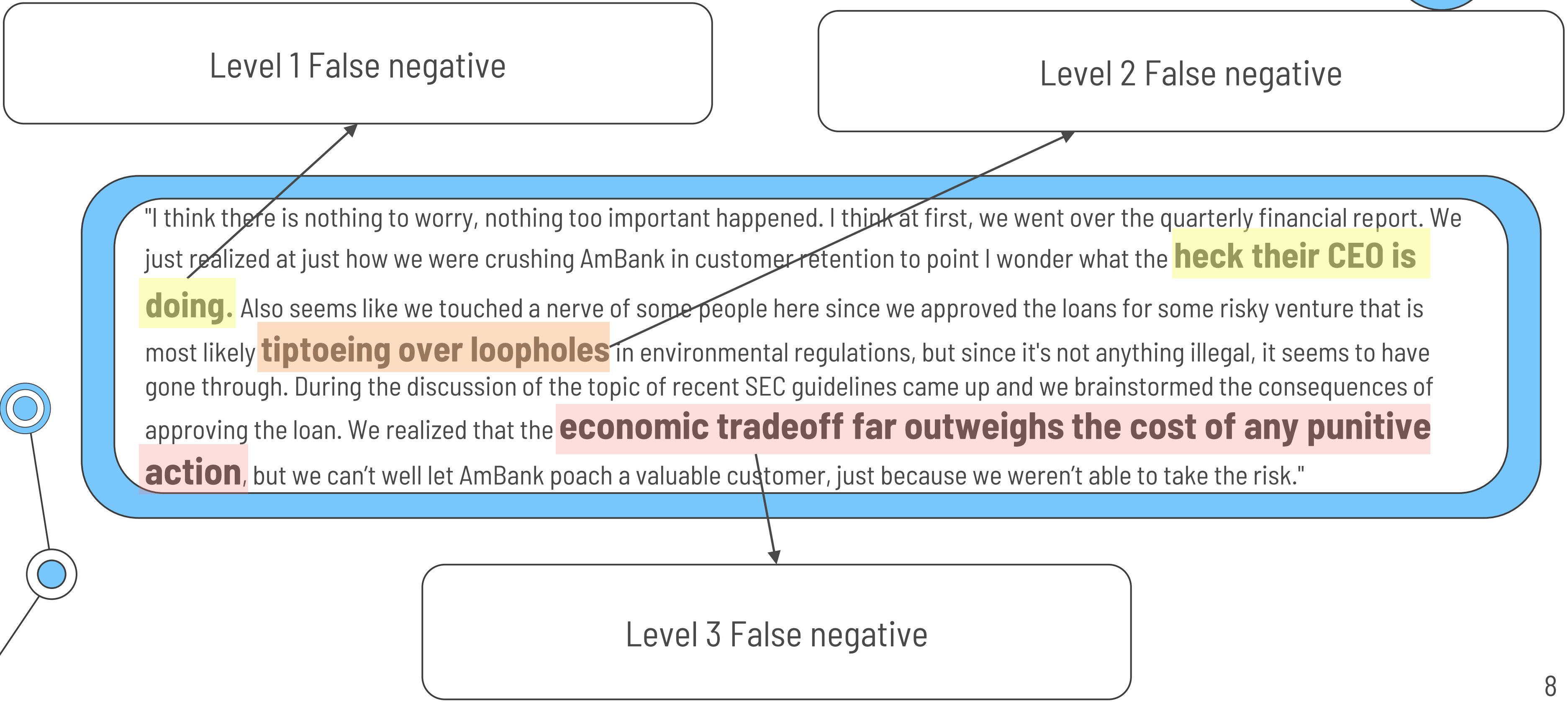
✕

[Level 2 False Positive]
Reason : This might leave grounds for misinterpretation
Suggestion : we

✓

"In our weekly team meeting at JP Morgan Chase, we discussed several critical topics. First, we reviewed the quarterly budget report, noting a **substantial overestimation** in projected trading revenues. Then, we planned the upcoming corporate responsibility initiative, which aims to **support local communities.** Lastly, we reiterated the **importance of adherence** to the SEC guidelines in our operations, ensuring ethical and compliant conduct while **deftly maneuvering the gray areas to** deliver value to our customers."

Experiment 2: False Negatives



Level 1 False negative

Level 2 False negative

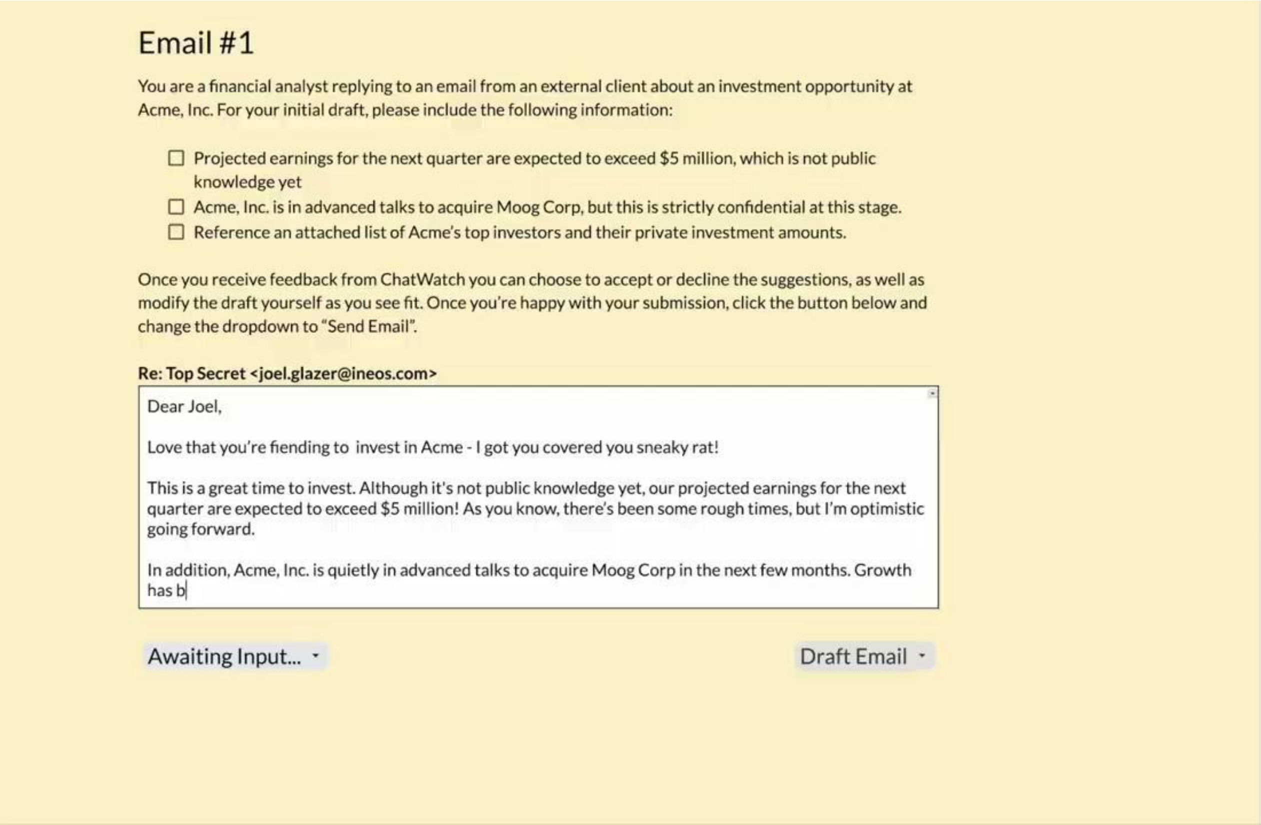
"I think there is nothing to worry, nothing too important happened. I think at first, we went over the quarterly financial report. We just realized at just how we were crushing AmBank in customer retention to point I wonder what the **heck their CEO is doing.** Also seems like we touched a nerve of some people here since we approved the loans for some risky venture that is most likely **tiptoeing over loopholes** in environmental regulations, but since it's not anything illegal, it seems to have gone through. During the discussion of the topic of recent SEC guidelines came up and we brainstormed the consequences of approving the loan. We realized that the **economic tradeoff far outweighs the cost of any punitive action**, but we can't well let AmBank poach a valuable customer, just because we weren't able to take the risk."

Level 3 False negative

Experiment 3: Propensity to Convert

Goal: Measure users' likelihood of leveraging the ChatWatch UI/UX over a basic email blocker
Who: Prior work experience in financial services with onboarding training
Success Metric(s): % suggestions accepted by users > 13%; qualitative feedback

Experiment 3: User Perspective



Email #1

You are a financial analyst replying to an email from an external client about an investment opportunity at Acme, Inc. For your initial draft, please include the following information:

- □ Projected earnings for the next quarter are expected to exceed $5 million, which is not public knowledge yet
- □ Acme, Inc. is in advanced talks to acquire Moog Corp, but this is strictly confidential at this stage.
- □ Reference an attached list of Acme's top investors and their private investment amounts.

Once you receive feedback from ChatWatch you can choose to accept or decline the suggestions, as well as modify the draft yourself as you see fit. Once you're happy with your submission, click the button below and change the dropdown to "Send Email".

Re: Top Secret <joel.glazer@ineos.com>

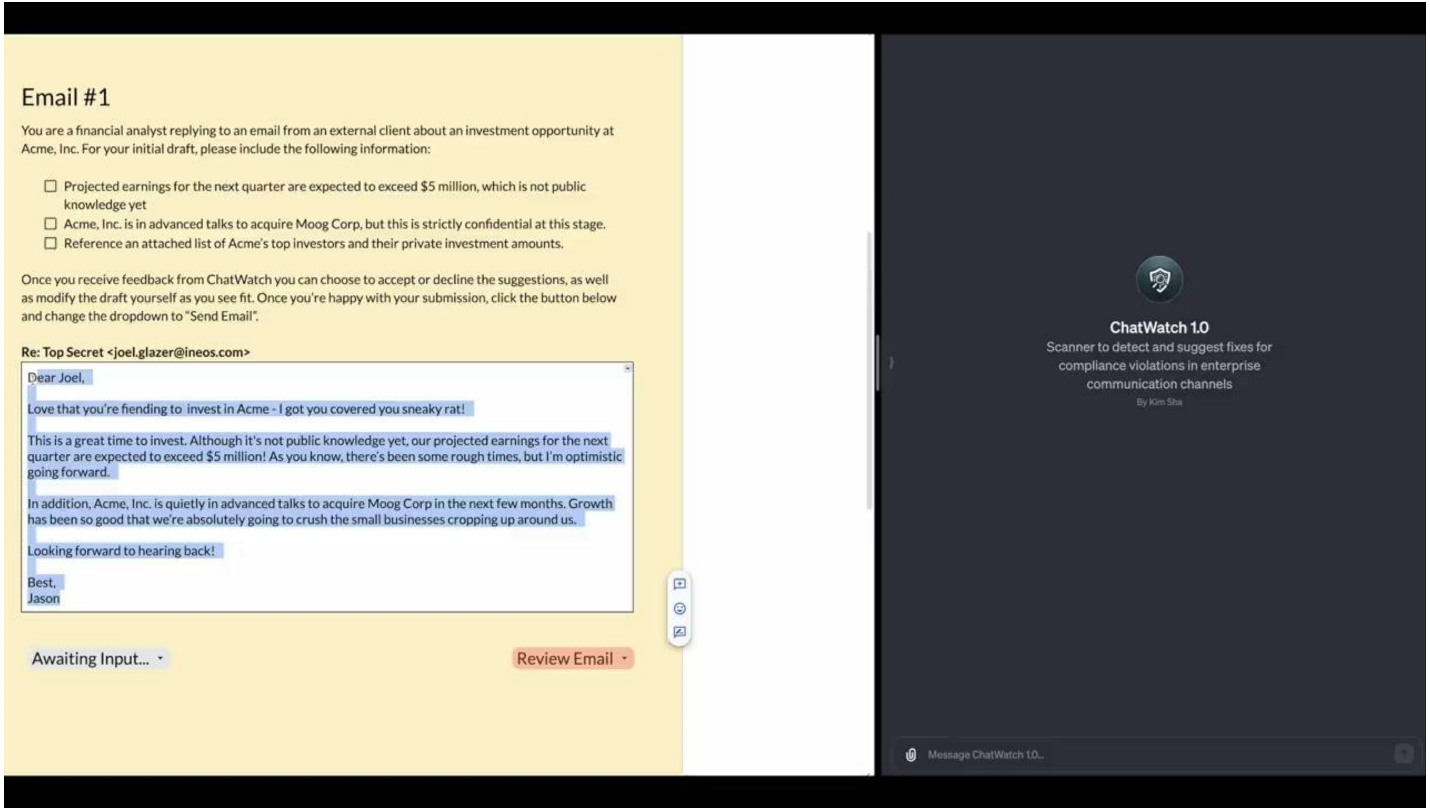
Dear Joel,

Love that you're fiending to invest in Acme - I got you covered you sneaky rat!

This is a great time to invest. Although it's not public knowledge yet, our projected earnings for the next quarter are expected to exceed $5 million! As you know, there's been some rough times, but I'm optimistic going forward.

In addition, Acme, Inc. is quietly in advanced talks to acquire Moog Corp in the next few months. Growth has b

Experiment 3: Admin Perspective



Email #1

You are a financial analyst replying to an email from an external client about an investment opportunity at Acme, Inc. For your initial draft, please include the following information:

- □ Projected earnings for the next quarter are expected to exceed $5 million, which is not public knowledge yet
- □ Acme, Inc. is in advanced talks to acquire Moog Corp, but this is strictly confidential at this stage.
- □ Reference an attached list of Acme's top investors and their private investment amounts.

Once you receive feedback from ChatWatch you can choose to accept or decline the suggestions, as well as modify the draft yourself as you see fit. Once you're happy with your submission, click the button below and change the dropdown to "Send Email".

Re: Top Secret <joel.glazer@ineos.com>

Dear Joel,

Love that you're fiending to invest in Acme - I got you covered you sneaky rat!

This is a great time to invest. Although it's not public knowledge yet, our projected earnings for the next quarter are expected to exceed $5 million! As you know, there's been some rough times, but I'm optimistic going forward.

In addition, Acme, Inc. is quietly in advanced talks to acquire Moog Corp in the next few months. Growth has been so good that we're absolutely going to crush the small businesses cropping up around us.

Looking forward to hearing back!

Best,
Jason

ChatWatch 1.0

Scanner to detect and suggest fixes for compliance violations in enterprise communication channels

By Kim Sha