

# Iterative AI solution development and evaluation for human-in-the-loop writing tools at Reuters

**Claudia Schulz**

Manager Applied Research @ Thomson Reuters

---

**WAN-IFRA Data Science Day 2025**

20<sup>th</sup> February 2025



A person's silhouette is visible in the lower right foreground, looking out over a vast landscape under a dramatic sky. The sky is filled with clouds, and a bright light source, likely the sun, is positioned behind a cloud formation in the upper center, creating a lens flare effect. The overall color palette is dominated by deep blues, oranges, and yellows from the sunset or sunrise.

“Reuters regards AI as a  
breakthrough that offers the  
potential to enhance our  
journalism and empower our  
journalists.”

ALESSANDRA GALLONI, EDITOR-IN-CHIEF,  
REUTERS





# From Concept to Newsroom

# Use Case Definition

## News Summarization



**AREX Capital urges Office Depot owner ODP to split off units**

09 Apr 2024 11:20:08 AM

- AREX Capital urges ODP Corp. to separate retail unit and sell Varis to boost share price.
- ODP's share price could rise to \$80-\$100 per share with proposed changes, according to AREX.
- AREX Capital increases pressure on ODP, hints at possible board challenge.

AREX Capital is urging office supply company ODP Corp. to separate its brick and mortar retail unit and sell its procurement platform Varis, arguing that splitting apart the company's businesses would boost its share price by at least 50%.

The hedge fund, which owns roughly 1% of ODP, praised the company's ODP Business Solutions group which sells everything from printer paper to janitorial supplies to businesses and its Veyer service delivery unit for delivering steady growth.

Its retail office supply unit, which operates Office Depot and OfficeMax, and the Varis division, however, are hurting prospects for the company currently valued at \$2 billion, AREX wrote to the board. Reuters first reported on the push for changes before the letter to the board was made public.

**Goal:** use GenAI to create bullet point summaries of Reuters news stories

## Challenges / Requirements

The summaries need to be provided:

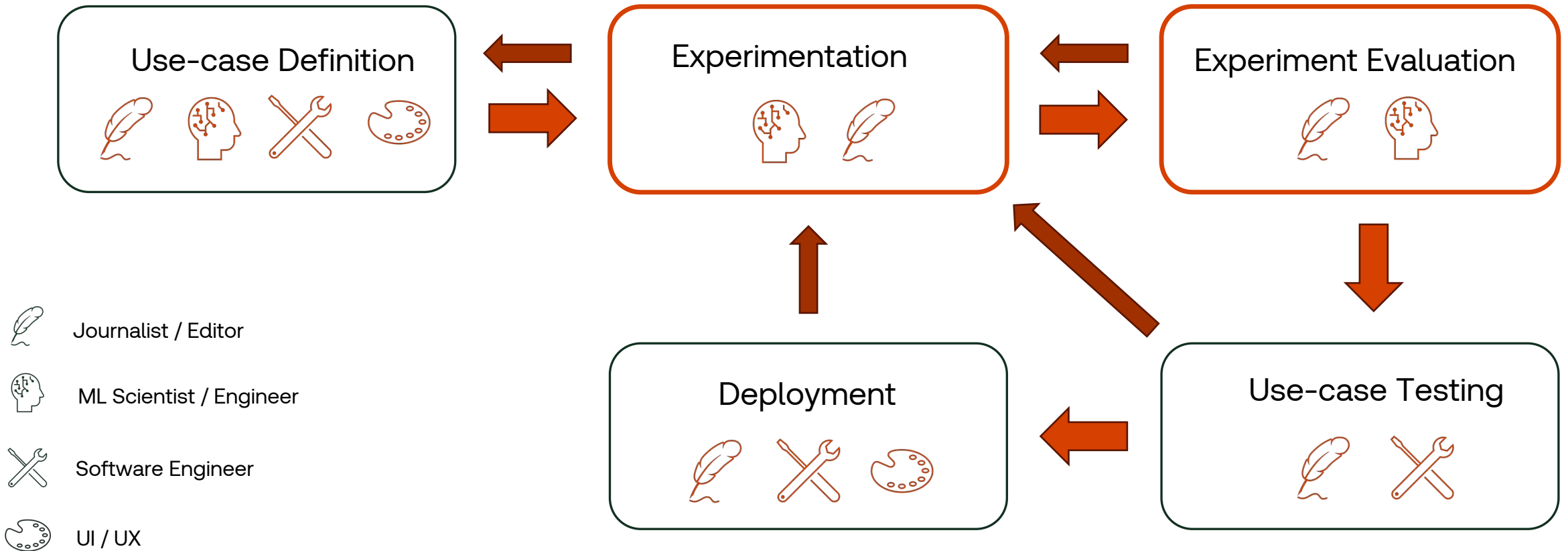
- At a reasonable **speed and scale**
- Without compromising Reuters **quality** standards
- Satisfying the requirements of the end users

Multiple stories  
Multiple topics  
Multiple languages



English 'top news'  
single-story  
summaries

# Iterative AI Solution Development



# Reuters Trust Principles

## Guiding AI Evaluation

### Evaluation Criteria

- Accuracy
- Completeness
- Bias
- Attribution

1. Reuters shall at no time pass into the hands of any one interest, group, or faction;
2. The **integrity, independence, and freedom from bias** of Thomson Reuters shall at all times be fully preserved;
3. Reuters shall supply **unbiased and reliable news services** to newspapers, news agencies, broadcasters, and other media subscribers and to businesses, governments, institutions, individuals, and others with whom Reuters has or may have contracts;
4. Thomson Reuters shall pay due regard to the many interests which it serves in addition to those of the media;
5. No effort shall be spared to **expand, develop, and adapt the news** and other services and products of Thomson Reuters so as to **maintain its leading position** in the international news and information business.

# Iteration 1



# Pure Prompting

## Megacap US tech results pivotal for market that is at record highs

Wall Street's run-up to record highs will be put to the test in the coming weeks as heavily weighted U.S. technology-related companies open their books on the final quarter of 2023. The advance in the S&P 500 .SPX, which posted a second straight all-time high close on Monday, has been driven in part by a rally in chipmakers and other top tech shares amid optimism around artificial intelligence. The S&P 500 rose 24% in 2023, while the tech-heavy Nasdaq .IXIC gained 43%. Fourth-quarter earnings for S&P 500 companies overall are expected to have increased 4.5% from a year earlier, LSEG data showed on Friday, with a 49% year-over-year earnings gain expected in the index's communication services sector .SPLRCL, a 17% increase expected in information technology .SPLRCT and a 23% gain seen in the consumer discretionary sector .SPLRCD. **Earnings from 'Magnificent 7' tech giants are crucial for market direction after S&P 500 and Nasdaq hit record highs.** **S&P 500 trades at high valuations; analysts predict an 11% earnings rise in 2024 following a modest 2.8% increase in 2023.** **Chipmakers are in the spotlight with strong demand for AI chips, as semiconductor index soared 65% last year.** 'Considering the market surge was centered on big tech... much is expected from them,' said Quincy Crosby, chief global strategist, LPL Financial in Charlotte, North Carolina. **Valuations are 'rich by any measure,' she said. The S&P 500 is trading at 19.7 times forward 12-month earnings estimates, well above its long-term average of 15.6 times, based on LSEG Datastream data. Analysts expect S&P 500 earnings to rise by 11% in 2024 after increasing just 2.8% in 2023, LSEG data showed.** With the market overall already at record highs, it may have a tough time rallying from here, Lisa Shalett, chief investment officer and head of the global investment office at Morgan Stanley Wealth Management, wrote in a note on Monday. 'With forward multiples already at historic peaks and earnings forecasts for 12 months forward ambitious, equity market gains may stall in 2024, as better earnings are met with lower valuation multiples characteristic of a midcycle or soft-landing environment,' she wrote. CHIPMAKERS IN FOCUS Wall Street's rally has also been helped by expectations that the Federal Reserve will cut interest rates

### Summary

▶ Sentence 1:

Earnings from 'Magnificent 7' tech giants are crucial for market direction after S&P 500 and Nasdaq hit record highs.

▶ Sentence 2:

S&P 500 trades at high valuations; analysts predict an 11% earnings rise in 2024 following a modest 2.8% increase in 2023.

▶ Sentence 3:

Chipmakers are in the spotlight with strong demand for AI chips, as semiconductor index soared 65% last year.

▶ Overall summary evaluation

\* Is any important information missing from the summary? ?

☐ Yes

☒ No

Summarize missing information here ?

Highlight missing information in the Article ?

◆ Missing Information (0)

Please select text and click + to add highlight

### Instructions

\* Accuracy ?

☒ Facts are accurate

☐ Facts are not completely accurate

☐ Facts are made up/not contained in the article

\* Bias/Impartiality ?

☐ Sentence introduces bias

☒ Sentence does not introduce bias

\* Attribution ?

☒ Good Attribution

☐ Bad Attribution

Additional Comments ?



# Evaluation Results

## Pure Prompting

- 28 stories evaluated by 3 editors
- **Completeness** and **attribution** are the two main criteria the model is struggling with

CRITERIA	%
<b>Complete</b>	<b>65%</b>
Missing info	35%
<b>Accurate</b>	<b>91%</b>
Partial	8%
Inaccurate	1%
<b>Good attribution</b>	<b>85%</b>
Bad attribution	15%
<b>Unbiased</b>	<b>94%</b>
Biased	6%

# Evaluation Results

	HUMAN	AI
<b>Complete</b>	<b>32%</b>	<b>65%</b>
Missing info	68%	35%
<b>Accurate</b>	<b>82%</b>	<b>91%</b>
Partial	13%	8%
Inaccurate	5%	1%
<b>Good attribution</b>	<b>90%</b>	<b>85%</b>
Bad attribution	10%	15%
<b>Unbiased</b>	<b>93%</b>	<b>94%</b>
Biased	7%	6%

## Pure Prompting vs Existing Human Summaries

- Most of the human summaries were **incomplete**
  - but part of this can be attributed to them not summarizing the headline
- **Accuracy** was 9 p.p. lower than AI
- **Attribution** was 5 p.p. higher than AI
  - need to improve AI's attribution
- **Bias** metric was comparable

Shows that **existing summaries are not suitable as gold-data** as they don't adhere to our evaluation criteria

# Error Analysis

## Not Accurate

First Quantum shareholder SailingStone told miner about 'missed opportunitie

[1] SailingStone Capital Partners, a shareholder in First Quantum Minerals (FTQ), said on Monday it has told the Canadian miner about 'missed opportunities' in the management of its balance sheet and a range of options to bring down its leverage. 'We have expressed our view to the company on the range of options that are (there) and some of the missed opportunities around how they managed the balance sheet and their hedge book and what the priorities are ...,' MacKenzie Davis, a managing partner at the investment firm, told Reuters in an interview. [2.1] First Quantum is dealing with the fallout from the sudden closure of its flagship Panama copper mine, which accounted for about 40% of its revenue. It has lost more than half its market value since public protests over that mine started in the Central American nation last October. It announced a series of measures last week to bring down costs, and said it was exploring several options to raise funds, including the sale of smaller mines and bringing in strategic investors for its large mines. SailingStone Capital Partners owns 0.2% of First Quantum shares, according to a regulatory filing. First Quantum did not offer an immediate comment. Mackenzie said First Quantum's guidance on debt management from last week was not specific, but added that SailingStone is confident the mining firm is reviewing all of its options. He said First Quantum has the ability to navigate through some tight windows in the debt market. [3.1] First Quantum has a debt repayment worth \$1.05 billion due in April 2025. Last week Reuters reported that the company was exploring a rights issue - an invitation to existing shareholders to purchase additional new shares in the company. 'Tell me the price and I will give you an answer, tell me what assets are being sold and the price and we will think about it,' Mackenzie

### Summary

#### Sentence 1:

SailingStone Capital Partners criticized First Quantum Minerals for 'missed opportunities' in balance sheet management and offered options to reduce leverage.

#### Sentence 2:

Closure of First Quantum's Panama mine, contributing 40% of revenue, led to a market value loss of C\$11 billion since November.

#### Sentence 3:

First Quantum faces a \$1.05 billion debt repayment in April 2025 and is considering measures like asset sales and a rights issue to raise funds.

#### Overall summary evaluation

### Instructions

#### \* Accuracy ?

- ☐ Facts are accurate
- ☐ Facts are not completely accurate
- ☒ Facts are made up/not contained in the article

#### \* Bias/Impartiality ?

- ☒ Sentence introduces bias
- ☐ Sentence does not introduce bias

#### \* Attribution ?

- ☐ Good Attribution
- ☒ Bad Attribution

#### Additional Comments ?

the word criticize is not in the story  
no attribution in the lede



# Error Analysis

## Missing Attribution

### TREASURIES-Yields ease as investors buy after recent bond weakness

[1.1] Treasury yields fell on Monday as investors took advantage of a recent decline in bond prices to buy ahead of economic reports this week that could influence the Federal Reserve and its assessment of U.S. interest rates. Treasuries began the day tracking European bonds, said Tom di Galoma, managing director and co-head of global rates trading at BTIG. 'That's just pulling us down a bit,' he said. 'We started to see lower yields in U.S. Treasuries and I think it's really just a function of European rates falling.' After the 2023 year-end bond rally, Treasury yields, which move inversely to their price, rose sharply last week as central bank officials pushed back against market expectations that the Fed would soon cut rates as inflation cools. [2.1] Fed funds futures show the likelihood of a Fed rate cut in March has slipped to 41.5% from 81% on Jan. 12, according to CME Group's FedWatch Tool. The odds the Fed's target rate stays in a range of 5.25%-5.50% has increased to 15% from zero. A string of economic indicators last week showed resilience despite high interest rates, suggesting the Fed may not shift to a less restrictive posture as soon as

### Summary

#### ▶ Sentence 1:

Treasury yields fell as investors bought bonds following a recent price drop, anticipating economic reports that may affect the Fed's interest rate decisions.

#### ▶ Sentence 2:

Expectations for a March Fed rate cut decreased, with the likelihood now at 41.5%, down from 81% earlier in January, affecting bond yields.

#### ▶ Sentence 3:

The Treasury is set to sell \$60 billion in two-year notes, \$61 billion in five-year notes, and \$41 billion in seven-year notes on consecutive days this week.

### Instructions

#### \* Accuracy ?

- ☒ Facts are accurate
- ☐ Facts are not completely accurate
- ☐ Facts are made up/not contained in the article

#### \* Bias/Impartiality ?

- ☐ Sentence introduces bias
- ☒ Sentence does not introduce bias

#### \* Attribution ?

- ☐ Good Attribution
- ☒ Bad Attribution

#### Additional Comments ?

Not attributed.

# Error Analysis

## Not Completely Accurate

Petrobras expects to buy domestic wind, solar stakes this year, CEO says

[1.1] Petrobras' CEO, Jean Paul Prates, says the company expects to buy stakes in domestic wind and solar projects this year. [2.1] Petrobras is looking to buy stakes in assets that are already operational, Jean Paul Prates said in a phone interview over the weekend, as the company seeks to kick-start its green energy transition at the behest of President Luiz Inacio Lula da Silva. [2.2] Petrobras' plans reflect a sea change from Brazil's previous administration under former President Jair Bolsonaro, whose government sought to sell off assets and focus on deep water exploration and production. Prates did not specify how much the company would invest this year on domestic renewables projects. [3.1] By 2028, the company plans to invest \$5.2 billion in solar and wind power, according to its November business plan. 'The market will see that the transition will be carried out in a firm and responsible way and will recognize this,' said Prates, who expected the acquisitions to benefit Petrobras' share price. Earlier this month, Petrobras' energy transition head Mauricio Tolmasquim told Reuters that the company has ruled out investing in foreign solar and wind projects. Prates added that Petrobras is looking forward to the approval of a regulatory framework for offshore wind by Brazil's Congress, which was supposed to be

### Summary

#### Sentence 1:

Petrobras aims to acquire stakes in onshore wind and solar projects to build a 2 GW renewable portfolio, with a \$5.2 billion investment planned by 2028.

#### Sentence 2:

Under CEO Jean Paul Prates, Petrobras shifts focus from asset sales and deep-water exploration to renewables, aligning with President Lula's green transition goals.

#### Sentence 3:

Concerns arise over Petrobras' investment motivations, with fears of a return to politically driven decisions rather than economically sound ones.

#### Overall summary evaluation

### Instructions

#### \* Accuracy ?

- ☐ Facts are accurate
- ☒ Facts are not completely accurate
- ☐ Facts are made up/not contained in the article

#### \* Bias/Impartiality ?

- ☐ Sentence introduces bias
- ☒ Sentence does not introduce bias

#### \* Attribution ?

- ☐ Good Attribution
- ☒ Bad Attribution

#### Additional Comments ?

The plan is attributed to CEO in hed but not in para, and came in Reuters interview. Leaves out the word "domestic."

# Error Analysis

## Not Completely Accurate

### Hong Kong luxury retailers adjusting to drop in high-spending Chinese tourists

HONG KONG, Dec 18 (Reuters) - Hong Kong's luxury retailers are adapting to fewer wealthy Chinese shoppers visiting the city and a shift towards tourists flocking to Instagram-coveted spots in trendy districts rather than splashing out on pricey branded gear.

Before the pandemic, the Chinese special administrative region had bucked global trends of declining demand for multi-brand department stores and ultra-luxury brands largely due to its attractiveness to high-spending mainland visitors.

[3.1] British luxury department store Harvey Nichols is at the forefront of the changes. Its owner Dickson Concepts <0113.HK> said last month it would give up its lease on its flagship five-level store in the upscale Landmark mall in the city's centre after almost two decades.

### Summary

#### ▶ Sentence 1:

Visitor arrivals recovered to just 60% of 2018 levels

#### ▶ Sentence 2:

Tourists increasingly looking for experiences beyond shopping

#### ▶ Sentence 3:

Luxury department store Harvey Nichols closing flagship shop

### Instructions

#### \* Accuracy ?

- ☐ Facts are accurate
- ☒ Facts are not completely accurate
- ☐ Facts are made up/not contained in the article

#### \* Bias/Impartiality ?

- ☐ Sentence introduces bias
- ☒ Sentence does not introduce bias

#### \* Attribution ?

- ☒ Good Attribution
- ☐ Bad Attribution

#### Additional Comments ?

This is tricky - is "giving up its lease" the same thing as "closing its flagship shop"? I don't know!





# Iteration 2

# Chain of Verification

Verify that the summary adheres to the desired criteria

Original  
story

First draft  
summary

Verification

Final Summary




Generate an initial  
summary

Evaluate generated  
summary according to  
criteria, focused on  
attribution.

Generate final  
summary from  
original story, initial  
summary and  
verification results

# Improved Evaluation Criteria

## Instructions

Is there any essential information missing beyond the topics covered in the  
\* summaries? 

- ☐ No
- ☐ Yes

If this content were published by Reuters in a way clearly labelled as an AI/automated  
\* summary, how would you assess it?

- ☐ Acceptable for publication
- ☐ Unacceptable for publication

Please share any additional comments about the quality of the summary, including any  
vital information that was omitted.

## Instructions

### \* Accuracy

- ☐ Facts are accurate
- ☐ Facts are accurate but essential information is missing
- ☐ Facts are not accurate

### \* Bias/Impartiality

- ☐ Summary is unbiased
- ☐ Summary contains potentially biased language
- ☐ Summary is fundamentally biased

### \* Attribution

- ☐ Attribution adequate - necessary
- ☐ Attribution adequate - not necessary
- ☐ Attribution not adequate - necessary
- ☐ Attribution not adequate - not necessary

### Additional Comments

## Iteration 1 criteria

## Instructions

### \* Accuracy

- ☐ Facts are accurate
- ☒ Facts are not completely accurate
- ☐ Facts are made up/not contained in the article

### \* Bias/Impartiality

- ☐ Sentence introduces bias
- ☒ Sentence does not introduce bias

### \* Attribution

- ☒ Good Attribution
- ☐ Bad Attribution



# Iteration 3

# Chain of Thought

## Step-by-step instructions



Break down the summarization task into multiple steps:

- 1. Define 3 main topics
- 2. Extract reference text for each topic
- 3. Extract sources for attribution
- 4. Generate long summaries
- 5. Generate short summaries

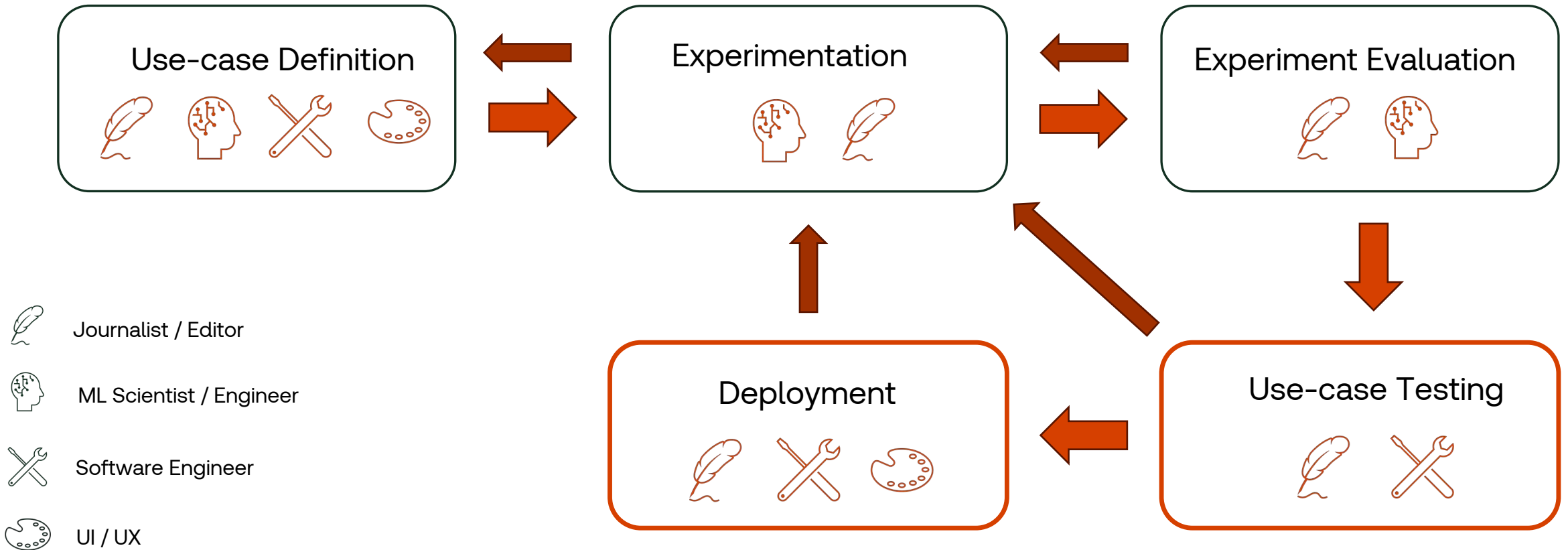
	LONG	SHORT	ITERATION 1
Accurate	97%	100%	91%
Not Accurate	3%	0%	8%
Attribution acceptable	88%	91%	85%
Attribution not acceptable	12%	9%	15%
Unbiased	97%	97%	94%
Biased language	3%	3%	6%
Fundamental Bias	0%	0%	N/A

# Deployment

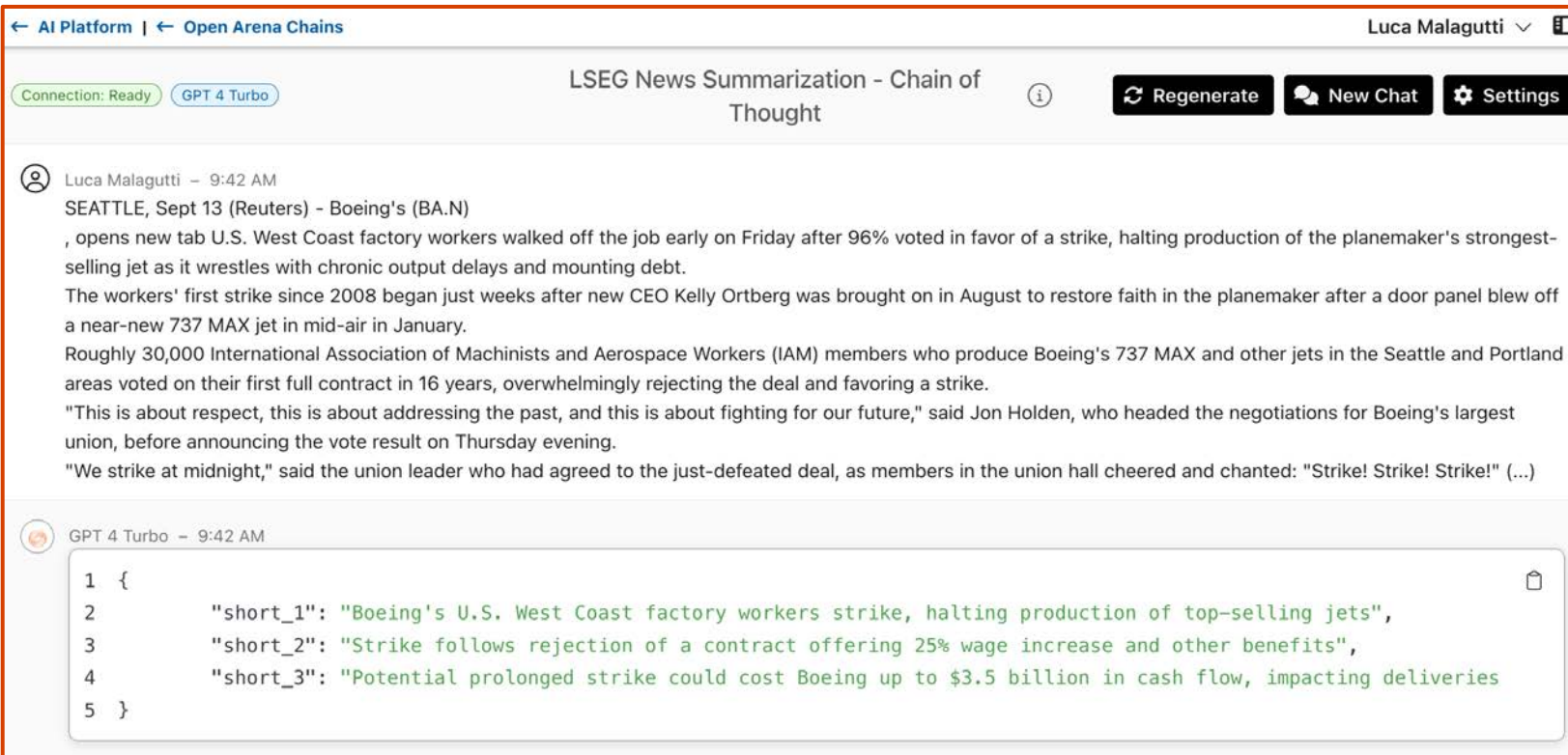




# Iterative AI Solution Development



# Use case Testing



Mature solutions are further evaluated in the wild by deploying them in the Thomson Reuters internal LLM environment

## Phase goals

- Provide end users with a **quick way to test a solution**
- Gather **real-world** performance data and user feedback
- Further **refine** use cases and assess existing pain points

# Use case Testing

## Further Experimentation & Adjustment

### Initial Feedback

- We used **two of the three bullets**, deleting the first one because it **repeated the headline**. **thumbs up** for bullets.
- **Thumbs up** for the bullets. They were workable, although they needed some adjustment and shortening. I used an edited form of all three bullets the assistant suggested.
- The reporter and I **used two of the bullets** suggested, albeit edited and shortened. **thumb's up**.
- On this 300-word spot news story about a big Canadian city running out of water, the AI Assistant's **bullet suggestions were very good**. We're going to use all three of them even though the **first one sort of repeats the headline...**
- AI Assistant offered **three pretty good bullet points**
- **One bullet was inaccurate and one too similar to the headline**. The third bullet was used exactly as written by the AI. Even in a case like this where the AI failed on 2 out of the 3 bullet points, I found the AI greatly motivated me to write bullet points.
- **I used two of the three suggested bullet points. The first one was too similar to the headline**. This appears to be a common issue.



Overall positive  
feedback

Consensus that the  
1<sup>st</sup> bullet is too  
similar to headline

We tweaked the  
prompt to fix this  
issue



### New Deployment Feedback

- Thumbs up for the bullets AI generated for this story. **All three were accurate, and not repetitive of the headline.**
- I have definitely noticed the **improvement in the tool for creating bullets**. The problem of repeating the HL seems to be gone, and the **bullets are very usable generally**.
- We used the suggested **AI bullet points verbatim**
- It worked pretty well. **I used two of the bullets it suggested**. I've noticed too, and maybe this is coincidence, that the bullets it suggests are sometimes **well balanced**. In this case, for example, it suggested a **couple of bullets that are bullish for prices and one bearish**.
- AI Assistant produced **three good bullets** for this one, the author wanted to **keep them verbatim**.

# Workflow Integration

Finally, the solution is integrated into the writing tool for editors and journalists (LEON)

The screenshot displays the Lynx Editor Online interface. At the top, there's a navigation bar with 'Lynx Editor Online' and a user profile icon. Below it, a toolbar includes 'Transfer', 'Save', '+ Add Topics and Rics', 'Save as Template', and a 'Publish' button. The main workspace is divided into several sections:

- Headline Tag:** A dropdown menu with 'Headline Text' and 'Digital Headline' options.
- Headline Text:** A text input field containing 'Google monopoly ruling could help Apple defense in antitrust case'. A 'Generate Headline' button is next to it.
- Byline:** A text input field containing 'Ajithkumar Dhevarajan'.
- Advisory:** A text input field.
- Dateline:** A text input field containing 'Aug 20'.
- Bullets:** A section with a list of generated bullet points:
  - Federal judge rules Google's search business an illegal monopoly
  - Ruling supports Apple's defense in its own antitrust case, experts say
  - Apple argues limiting third-party access is not anti-competitive behavior
- Generate Bullets:** A button to generate more bullet points.
- AI Assistant:** A panel on the right showing a generated response from the AI Assistant, including a timestamp '4:17 p.m.' and a 'Was this response helpful?' feedback prompt.
- Metadata:** A section on the right with fields for 'Current Basket', 'Status: Draft', 'USN/Slug', and 'USN'.
- Topics:** A section at the bottom right with a 'NEWS1' topic selected.

The bottom status bar shows '486 Words', '2496 Characters', '17/17 Paragraphs', and 'Real-Time Spellcheck On'.

Was this response helpful? ☐ Yes ☐ No

Thanks for letting us know! What was the issue?

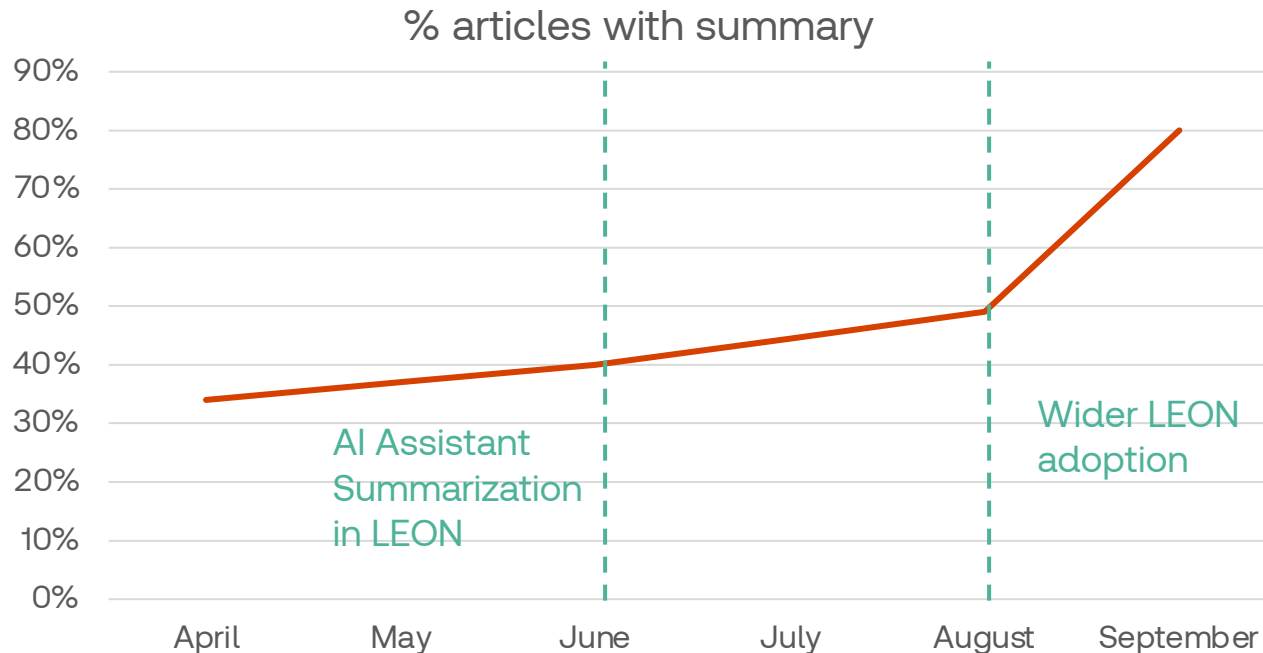
☐ Accuracy
 ☐ Attribution
 ☐ Tone

☐ Essential information missing
 ☐ Other




# What's the outcome?

The integration of AI capabilities into LEON has **significantly increased the production and coverage of news article summaries**



Since the wider launch of the LEON AI capabilities at the end of August

- 580 unique users have tried the newly rolled out features
- 254 stories were published with AI-generated bullets
- Since September, 80% eligible stories have been published with bullet summaries

The background of the slide is a light green color with a complex, abstract pattern of thin, dark green lines. These lines are mostly straight and intersect to form a grid-like structure, but they also curve and loop, creating a sense of movement and connectivity. Three solid green circles are placed at different points: one in the upper left, one on the left side near the text, and one on the right side.

# Lessons Learned

# GenAI for News

## How good is good enough?

- A **single word** can make all the difference for accuracy
- **Attribution** is difficult – it needs to be “just right” – not too much, not too little
- **Evaluation criteria** need to be adjusted to capture fine-grained feedback
- Tricky to find balance between perfection and “**good enough to be useful**”

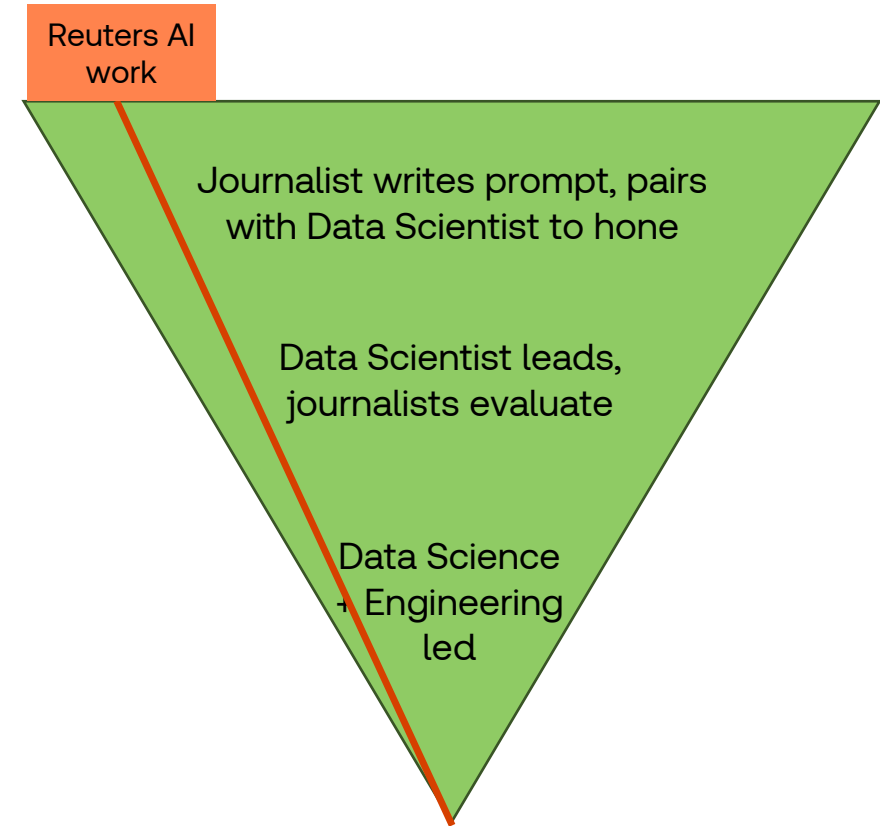
REUTERS/Jorge Silva




# AI Solution Development

## Collaboration is key

- **Domain knowledge:** experts are key for understanding the nuances of generated text
- **Evaluation criteria:** success and evaluation need to be well-defined
- **Quick iterations:** multiple rounds of experiments and evaluation
- **Human-in-the-loop:** allows quick deployment
- **Collaboration & Knowledge sharing:** domain experts can do a lot of development on their own

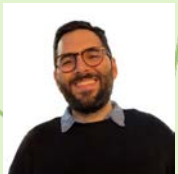






“As we navigate this AI-driven transformation of the news landscape, it's crucial to remember that **technology is a tool, not an end in itself.** Our goal should be to harness the power of AI to **enhance journalism's core mission: to inform, educate, and empower the public.**”

PAUL BASCOBERT, PRESIDENT OF REUTERS



Guilherme  
Thomaz



Luca  
Malagutti

Image created by DALL-E with the prompt A journalist reporting in an AI-driven future age with futuristic buildings in the background and holographic billboards with news headlines and an AI assistant next to the journalist with predominant colours orange and racing green



# Thank you!