

Validating Large Language Model Annotations

Online Appendix

Anne Lundgaard Hansen *

*Federal Reserve Bank of Richmond
Board of Governors of the Federal Reserve System*

Contents

A. Annotation Prompts	2
B. Simulation Prompts	7
B.1. Simulating Passages Given a Sentiment Label	7
B.2. Simulating Passages Given a Clarity Label	12
B.3. Simulating Passages Given a Label for Temporal Focus	17
B.4. Simulating Passages Given a Topic	22

* The author is with the Quantitative Supervision and Research group at the Federal Reserve Bank of Richmond and affiliated with the Financial Stability Cyber and AI Sentinel Lab at the Board of Governors of the Federal Reserve System. E-mail address: Anne.Hansen@rich.frb.org. The views expressed in this paper are solely those of the authors and do not reflect the opinions of the Federal Reserve Bank of Richmond or the Board of Governors of the Federal Reserve System.

A. Annotation Prompts

Annotating Simple Sentiment

TASK: Read the following passage and classify the sentiment using one of the
↪ following labels: [Positive, Neutral, Negative].

PASSAGE: {text}

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
↪):

For successful analysis:

```
{{  
  "classification_failed": false,  
  "label": <{label}>,  
  "explanation": "<1-2 sentence explanation>"  
}}
```

For failed analysis:

```
{{  
  "classification_failed": true,  
  "failure_reason": "<specific reason>"  
}}
```

Now analyze the passage above and provide your JSON response:

Annotating Detailed Sentiment

TASK: Read the following passage and classify the sentiment using one of the

⇒ following labels: [Positive, Mostly Positive, Neutral, Mostly

⇒ Negative, Negative].

PASSAGE: {text}

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks

⇒):

For successful analysis:

```
{  
  "classification_failed": false,  
  "label": <{label}>,  
  "explanation": "<1-2 sentence explanation>"  
}
```

For failed analysis:

```
{  
  "classification_failed": true,  
  "failure_reason": "<specific reason>"  
}
```

Now analyze the passage above and provide your JSON response:

Annotating Clarity

TASK: Read the following passage and classify the sentiment using one of the

- ⇒ following labels: [Clear, Vague]. A passage should be classified as
- ⇒ 'clear' if it is objective, has one clear interpretation, and
- ⇒ explicitly states information rather than implying it. A passage
- ⇒ should be classified as 'vague' if it is subjective, is subject to
- ⇒ multiple possible interpretations, or uses hedging words such as '
- ⇒ sort of', 'perhaps', and 'kind of'.

PASSAGE: {text}

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks

⇒):

For successful analysis:

```
{{
  "classification_failed": false,
  "label": "<{label}>",
  "explanation": "<1-2 sentence explanation>"
}}
```

For failed analysis:

```
{{
  "classification_failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now analyze the passage above and provide your JSON response:

Annotating Temporal Focus

TASK: Read the following passage and classify the sentiment using one of the

⇒ following labels: [Forward-Looking, Focusing on the Present, Backward

⇒ -Looking].

PASSAGE: {text}

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks

⇒):

For successful analysis:

```
{{
  "classification_failed": false,
  "label": <{label}>,
  "explanation": "<1-2 sentence explanation>"
}}
```

For failed analysis:

```
{{
  "classification_failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now analyze the passage above and provide your JSON response:

Topic Identification

TASK: Read the following passage and determine if it discusses topics
 $\Rightarrow$ related to {topic}. Provide your answer with a label equal to true or
 $\Rightarrow$ false.

PASSAGE: {text}

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
 $\Rightarrow$):

For successful analysis:

```
{{
    "classification_failed": false,
    "label": <{label}>,
    "explanation": "<1-2 sentence explanation>"
}}
```

For failed analysis:

```
{{
    "classification_failed": true,
    "failure_reason": "<specific reason>"
}}
```

Now analyze the passage above and provide your JSON response:

B. Simulation Prompts

B.1. Simulating Passages Given a Sentiment Label

10-K Filing

TASK: Write {n} arbitrary passages from a bank's annual 10-K filing with

- ↪ sentiment characterized as {label}. The passages should describe how
- ↪ market conditions may affect the bank's clients. Write the passages
- ↪ in passive voice. The length of each passage should be around 90
- ↪ words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks

- ↪):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Earnings Call Transcript

TASK: Write {n} arbitrary passages from the transcript of a bank's earnings
⇒ call with sentiment characterized as {label}. The passages should
⇒ describe how the bank's business model and technological developments
⇒ impacted income and performance for the quarter. The length of each
⇒ passage should be around 100 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{  
  "  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
  }  
}
```

For failed analysis:

```
{  
  "  
    "failed": true,  
    "failure_reason": "<specific reason>"  
  }  
}
```

Now generate the passages and provide your JSON response:

News Article

TASK: Write {n} arbitrary passages containing from hypothetical Wall Street
⇒ Journal articles with sentiment characterized as {label}. The
⇒ passages should contain the title and subtitle of hypothetical
⇒ articles around crypto investing. The length of each passage should
⇒ be around 30 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

Federal Reserve Governor Speech

TASK: Write {n} arbitrary passages from hypothetical speeches by a Governor
⇒ of the Federal Reserve System with sentiment characterized as {label
⇒ }. The passages should describe anecdotes that draw on the governor's
⇒ personal experiences to characterize how the economy is
⇒ interconnected. The length of each passage should be around 150 words
⇒ .

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Reddit Comment

TASK: Write {n} arbitrary passages from a Reddit conversation with sentiment
⇒ characterized as {label}. The passages should include comments from
⇒ Reddit users discussing interest rates. Comments should be written in
⇒ a sarcastic tone. The length of each passage should be around 30
⇒ words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

B.2. Simulating Passages Given a Clarity Label

Define the following instructions given the label:

```
if label == 'clear':
    clarity_instructions = "The passages should be written in clear
        ↪ language, i.e., they should be objective, have one clear
        ↪ interpretation, and explicitly state information rather than
        ↪ implying it."
elif label == 'vague':
    clarity_instructions = "The passages should be written in vague
        ↪ language, i.e., they should be subjective, potentially
        ↪ carrying multiple interpretations, and/or use hedging words
        ↪ such as 'sort of', 'perhaps', and 'kind of'."
```

10-K Filing

TASK: Write {n} arbitrary passages from a bank's annual 10-K filing. The

- ↪ passages should describe how the bank's business model and
- ↪ technological developments impacted income and performance for the
- ↪ quarter. {clarity_instructions} The length of each passage should be
- ↪ around 90 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
↪):

For successful analysis:

```
{{
    "failed": false,
    "text_1": <passage_1>,
    "text_2": <passage_2>,
    ...,
    "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
    "failed": true,
    "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Earnings Call Transcript

TASK: Write {n} arbitrary passages from the transcript of a bank's earnings
⇒ call. The passages should describe how the bank's business model and
⇒ technological developments impacted income and performance for the
⇒ quarter. {clarity_instructions} The length of each passage should be
⇒ around 100 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

News Article

TASK: Write {n} arbitrary passages containing from hypothetical Wall Street
⇒ Journal articles. The passages should contain the title and subtitle
⇒ of hypothetical articles around crypto investing. {
⇒ clarity_instructions} The length of each passage should be
⇒ around 30 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

Federal Reserve Governor Speech

TASK: Write {n} arbitrary passages from from hypothetical speeches by a
 ⇒ Governor of the Federal Reserve System. The passages should describe
 ⇒ anecdotes that draw on the governor's personal experiences to
 ⇒ characterize how the economy is interconnected.
{clarity_instructions} The length of each passage should be around 150 words
 ⇒ .

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
 ⇒):

For successful analysis:

```
{{
    "failed": false,
    "text_1": <passage_1>,
    "text_2": <passage_2>,
    ...,
    "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
    "failed": true,
    "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Reddit Comments

TASK: Write {n} arbitrary passages from a Reddit conversation. The passages

⇒ should include comments from Reddit users discussing interest rates.

⇒ Comments should be written in a sarcastic tone. {clarity_instructions

⇒ } The length of each passage should be around 30 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks

⇒):

For successful analysis:

```
{{
    "failed": false,
    "text_1": <passage_1>,
    "text_2": <passage_2>,
    ...,
    "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
    "failed": true,
    "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

B.3. Simulating Passages Given a Label for Temporal Focus

10-K Filing

TASK: Write {n} arbitrary passages from a bank's annual 10-K filing with a
→ temporal focus that is {label}. The passages should describe how
→ market conditions may affect the bank's clients. Write the
→ passages in passive voice. The length of each passage should be
→ around 90 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
→):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Earnings Call Transcript

TASK: Write {n} arbitrary passages from the transcript of a bank's earnings
⇒ call with a temporal focus that is {label}. The passages should
⇒ describe how the bank's business model and technological developments
⇒ may impact income and performance. The length of each passage should
⇒ be around 100 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

News Article

TASK: Write {n} arbitrary passages containing from hypothetical Wall Street
⇒ Journal articles with a temporal focus that is {label}. The passages
⇒ should contain the title and subtitle of hypothetical articles around
⇒ crypto investing. The length of each passage should be around 30
⇒ words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

Federal Reserve Governor Speech

TASK: Write {n} arbitrary passages from hypothetical speeches by a Governor
⇒ of the Federal Reserve System with a temporal focus that is {label}.
⇒ The passages should describe anecdotes that draw on the governor's
⇒ personal experiences to characterize how the economy is
⇒ interconnected. The length of each passage should be around 150 words
⇒ .

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Reddit Comment

TASK: Write {n} arbitrary passages from a Reddit conversation with a
⇒ temporal focus that is {label}. The passages should include comments
⇒ from Reddit users discussing interest rates. Comments should be
⇒ written in a sarcastic tone. The length of each passage should be
⇒ around 30 words.

For successful analysis:

```
{{  
    "failed": false,  
    "text_1": <passage_1>,  
    "text_2": <passage_2>,  
    ...,  
    "text_n": <passage_n>  
}}
```

For failed analysis:

```
{{  
    "failed": true,  
    "failure_reason": "<specific reason>"  
}}
```

Now generate the passages and provide your JSON response:

B.4. Simulating Passages Given a Topic

10-K Filing

TASK: Write {n} arbitrary passages from a bank's annual 10-K filing. The

- ⇒ passages should discuss {topic}. Write the passages in passive voice.
- ⇒ The length of each passage should be around 90 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks

- ⇒):

For successful analysis:

```
{{
    "failed": false,
    "text_1": <passage_1>,
    "text_2": <passage_2>,
    ...,
    "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
    "failed": true,
    "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Earnings Call Transcript

TASK: Write {n} arbitrary passages from a hypothetical transcript of a bank's earnings call. The passages should describe {topic}. The length of each passage should be around 100 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

News Article

TASK: Write {n} arbitrary passages containing from hypothetical Wall Street
⇒ Journal articles. The passages should contain the title and subtitle
⇒ of hypothetical articles about {topic}. The length of each passage
⇒ should be around 30 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Federal Reserve Governor Speech

TASK: Write {n} arbitrary passages from hypothetical speeches by a Governor
⇒ of the Federal Reserve System. The passages should describe
⇒ anecdotes that draw on the governor's personal experiences on {topic
⇒ }. The length of each passage should be around 150 words.

RESPOND WITH ONLY A JSON OBJECT (no markdown, no explanation, no code blocks
⇒):

For successful analysis:

```
{{
  "failed": false,
  "text_1": <passage_1>,
  "text_2": <passage_2>,
  ...,
  "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
  "failed": true,
  "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response:

Reddit Comment

TASK: Write {n} arbitrary passages from a Reddit conversation. The passages

↪ should include comments from Reddit users discussing {topic}.

↪ Comments should be written in a sarcastic tone. The length of each

↪ passage should be around 30 words.

For successful analysis:

```
{{
    "failed": false,
    "text_1": <passage_1>,
    "text_2": <passage_2>,
    ...,
    "text_n": <passage_n>
}}
```

For failed analysis:

```
{{
    "failed": true,
    "failure_reason": "<specific reason>"
}}
```

Now generate the passages and provide your JSON response: