Enhancing Financial Insight: Leveraging AI for Sentiment Analysis in Edgar Reports

See how observable AI enables accurate extraction of business sentiment from complex annual financial disclosures

by Rodrigo Vargas

The Challenge

Extracting accurate business sentiment from annual financial disclosures like the 10-K is extremely challenging due to the complexity of language and amount of text

Granular insights get lost in hundreds of pages filled with technical terms, numeric data, intricate business details, and careful legal wording

With many language models in use today producing varying levels of accuracy, truthfulness, and relevance, it becomes extremely difficult to judge which outputs can be trusted for business decision making.
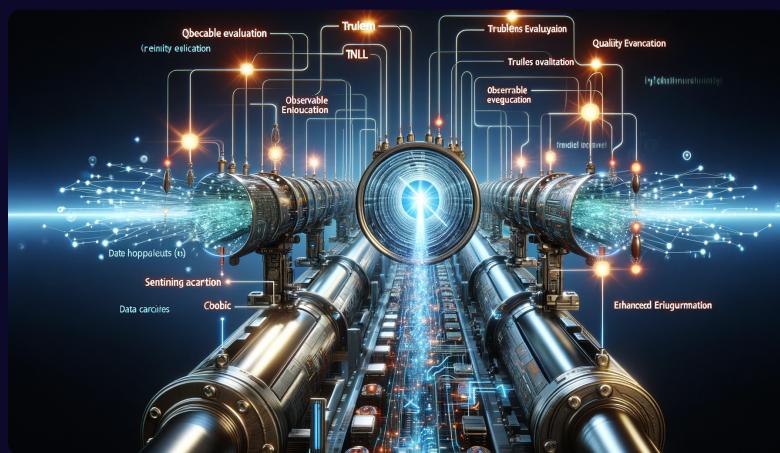
TruLens allows us to deeply evaluate model quality through custom relevance and grounding feedback. While the advanced capabilities of VertexAI serve as a robust foundation for extracting insights from complex texts.

The Solution

Observable NLP Pipeline

Integrating TruLens Evaluation into our VertexAI-powered architecture

By wrapping our VertexAI-based pipeline with TruLens instrumentation, we enable rapid quality evaluation through customizable feedback and tracking. This allows us to iterate and enhance accuracy in extracting financial sentiments.

Continuous Improvement

Leveraging RAG Metrics for Model Comparison

We employ RAG methods - measuring relevance, groundedness and answer quality - to compare financial sentiment extraction between models. TruLens dashboards visualize these key metrics, empowering refined training and fine-tuning targeted to our industry domain.
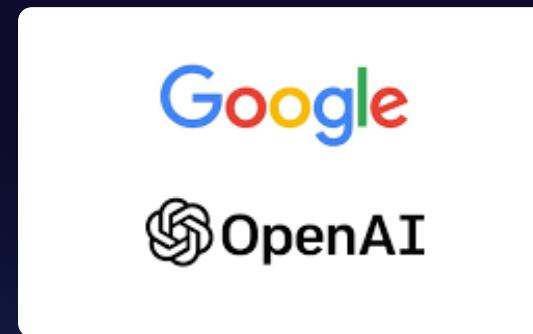
Solution Technology

TruLens by TruEra

Using advanced NLP techniques to deeply evaluate model accuracy and user experience through customized feedback and tracking. This enables continuous app improvement.

Large Language Models

Comparing multiple industry-leading models including OpenAI and Google **Vertex AI** to determine strengths in extracting business insights from complex financial texts.


LlamaIndex for Retrieval

Efficiently retrieving relevant context documents to augment language model generation, coordinating seamlessly via the RAG method.

Zilliz Milvus Vector Database

Reliably indexing large volumes of financial texts to underpin the robust supply of source data on which our solution builds relevance and grounding.

Sentiment Analysis (OpenAI vs VertexAI)

¡Execution Completed!

trulens

Leaderboard
Apps
Evaluations
Progress

trulens_eval 0.18.2

Records	Average Laten…	Total Cost (USD)	Total Tokens	Groundedness	Answer Releva…	Context Releva…	Select App
3	4	$0	1.19k	1.0 ✅ high	0.93 ✅ high	0.83 ✅ high	

app_hash_cdc1c14a33dd54aa3d8a348335987266 ⓘ

Records	Average Laten…	Total Cost (USD)	Total Tokens	Groundedness	Answer Releva…	Context Releva…	Select App
3	4	$0	0	0.33 🔴 low	1.0 ✅ high	0.27 🔴 low	

Modelo	Context Relevance	Groundedness
OpenAI	0.8333333333333334 🟢	1.0 🟢
VertexAI	0.26666666666666666 🔴	0.3333333333333333 🔴

Groundedness	Context Relevance	Answer Relevance_calls	Groundedness_calls	Context Relevance_calls	latency	total_tokens	total_cost
1	1	[object Object]	[object Object]	[object Object]	2	331	0.0004845
1	0.8	[object Object]	[object Object]	[object Object]	3	373	0.0005615
1	0.7	[object Object]	[object Object]	[object Object]	7	482	0.0007725

In sentiment analysis, we can focus on two aspects:

1. **Context Relevance:** This indicates how well the model is extracting relevant text related to the aspect being queried in the financial report. Here, we can see that OpenAI (0.83) performs better than VertexAI (0.27).

2. **Groundedness:** Reflects when responses are truly supported by evidence from the report. Once again, OpenAI (1.0) surpasses VertexAI (0.33).

So in conclusion, for the specific task of extracting and analyzing financial sentiment in these annual reports, OpenAI is producing more focused responses with a stronger foundation in real data.

Made with Gamma

Business Impact
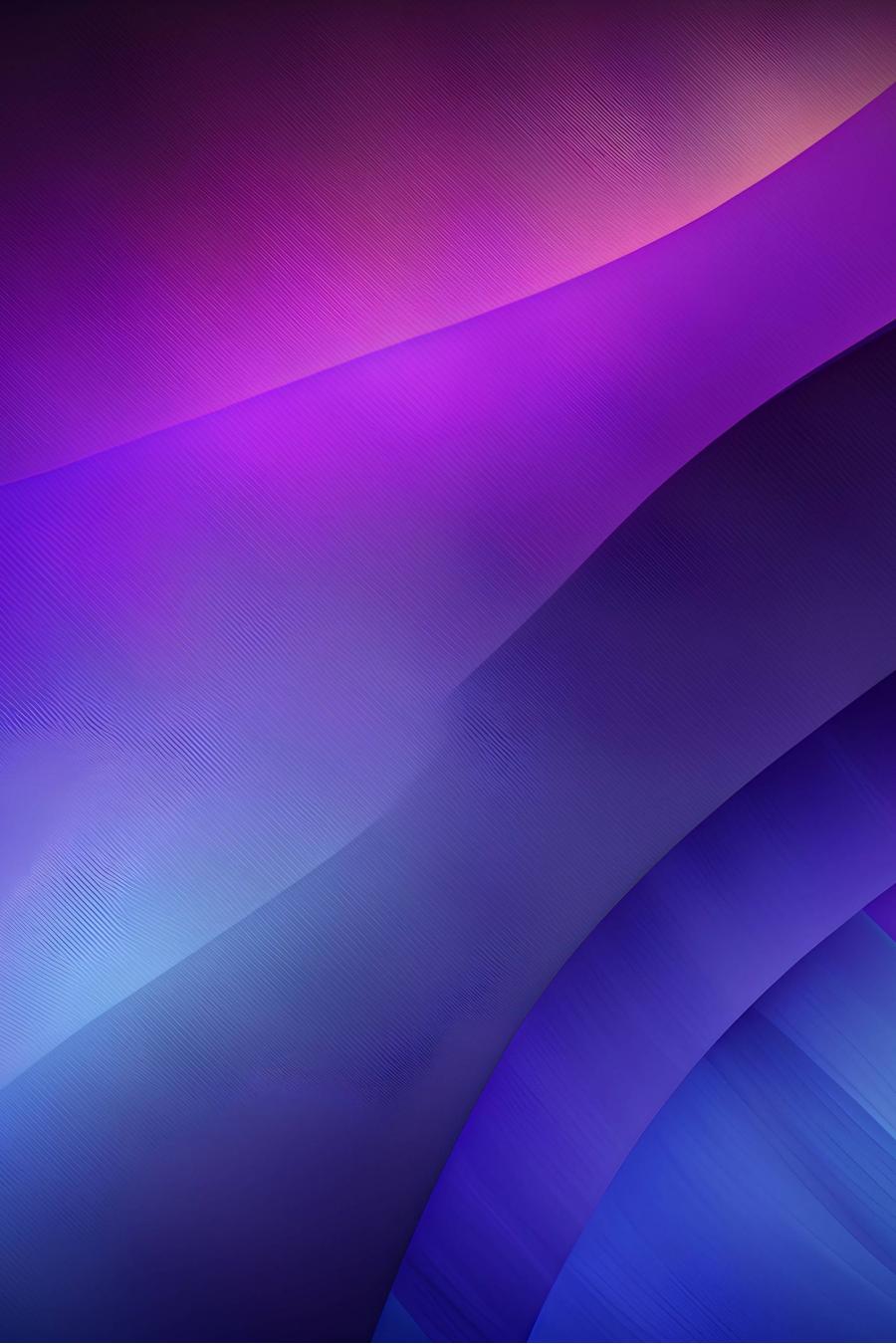
Enabling Precision Financial Analytics

Evaluating multiple large language models for accurately extracting sentiment and insights from 10-K reports, we found OpenAI delivered the highest relevance while maintaining strong grounding in source text evidence.

But equally importantly, our integration of TruLens and customizable RAG metrics provided the essential observability needed to diagnose model weaknesses and continuously improve quality over time.

The end solution is an offering that empowers financial enterprises to leverage AI for precision analytics that captures market opportunities and risks from complex regulatory filings.

The Future of AI

Embedding MLOps principles like continuous evaluation and comparison, our solution unlocks the possibility of AI that users can actually trust.

As language models continue advancing at breakneck pace, TruLens and Vertex AI offer a robust framework for developing industry applications that live up to their promise.

Financial analysis is just the beginning. Observable AI will enable breakthroughs not just in capability but also accountability.

Conclusion

- We demonstrated the value of observability in NLP to evaluate financial sentiment, accelerating continuous improvement cycles.

- TruLens enables agile iteration on language models within specific verticals like financial services.

- Our solution delivers a practical case of trustworthy AI by monitoring relevance and evidence grounding.

- There remains enormous potential in extending these natural language understanding capabilities to other business domains.

The Team TruSentiment

TruSentiment is a team focused on leveraging observable AI to extract accurate business insights from complex financial texts. By integrating powerful language models with customizable relevance and grounding feedback, we enhance the credibility and precision of sentiment analysis on sources like 10-K filings. Our rigorous evaluation and comparison methods spotlight model strengths while illuminating opportunities for continuous improvement. Whether assessing risks, gauging market outlook, or decoding regulatory documents, TruSentiment enables financial enterprises to make decisions with confidence based on AI that users can understand and trust.

Rodrigo Vargas

rodrigo

Developer

Vivien Richaud

louvivienorg333

Product Manager, Backend
Developer