

# MIDS DATASCI 210 Summer 2024 Section 1

# DIVITIAE.

—Empowering Your Investment Research with Artificial Intelligence—

David Campos | Krystyna Perez | Andrew Abrahamian | Dean Nakada | Geoffrey Wilson  
Date: August 5, 2024

# Team DIVITIAE.



David Campos  
Lead



Krystyna Perez  
AI



Andrew Abrahamian  
Quant



Dean Nakada  
Marketing



Geoffrey Wilson  
Legal & Sales

# Agenda

1. Problem and Customer Space
2. Solution Space
3. Overall Architecture
4. Technology: Quantitative Processing
5. Technology: Generative AI
6. Looking Forward

01

# Problem and Customer Space

“As a **retail investor**, I am **wasting time** researching investment opportunities across disparate sources. I need a streamlined source of intelligence.”

# Competition



yahoo/finance Search for news, symbols or companies

My Portfolio News Markets Sectors Screeners Personal Finance Videos

Summary  
News  
Chart  
Community  
Statistics  
Historical Data  
Profile  
Financials  
Analysis  
Options  
Holders  
Sustainability

NasdaqGS - Nasdaq Real Time Price - USD  
**NVIDIA Corporation (NVDA)** ☆ Follow ↕ Compare

**126.40 +0.31 (+0.25%)** **1.94%**  
At close: 4:00 PM EDT

**Valuation Measures**

Annual Monthly

	Current	1/31/2024	10/31/2023	7/31/2023	4/30/2023
Market Cap	3.11T	3.11T	1.01T	1.15T	686.27B
Enterprise Value	3.00T	3.00T	1.01T	1.15T	685.00B
Trailing P/E	72.34	72.34	31.09	31.09	31.09
Forward P/E	48.54	35.71	31.09	31.09	31.09
PEG Ratio (5yr expected)	1.45	1.16	0.60	0.60	0.60
Price/Sales	39.50	35.37	34.14	31.09	45.06
Price/Book	63.26	49.45	45.57	36.57	47.05
Enterprise Value/Revenue	38.71	34.64	33.63	30.62	44.46
Enterprise Value/EBITDA	60.94	59.31	66.16	77.37	178.68
				114.42	

Too many tabs / metrics to find information

Not all information is freely available.

Advanced Search

NVIDIA Past 24 hours

All Symbols People Pages **Headlines**

All Analysis News Transcripts Education

**The Rise Of Nvidia**  
Ironman at Political Calculations • Today, 10:15 AM • 21 Comments  
... stories of all time. The soaring stock price of **Nvidia** (NASDAQ: NVDA) has presented in the chart. Since that date, **Nvidia's** stock price has experienced considerable volatility.

**Nvidia** ticks up as Citi bumps price target on GB200 plans  
Chris Ciacca, SA News Editor • Today, 7:26 AM • 33 Comments  
... **Nvidia** (NVDA) was in focus on Wednesday as Citi raised its price target on the semiconductor. 40% of **Nvidia's** data center sales, enterprise AI agents are the ... Malk said, Malk maintained his Buy rating on **Nvidia** from \$126. ...

**NVD: How To Short Nvidia In Your IRA**  
Paul Franke • Today, 10:31 AM • 21 Comments  
... Alpha Table - **Nvidia**. "F" Quant Valuation Grade, June 24th, 2024. **Nvidia** drop. YCharts - **NVD** vs **NVIDIA**, Total Returns, June 1st, 2024. **NVD** vs **NVIDIA**, 1-Month Total Returns ...

**Nvidia's** market cap is down more than \$400B in just 3 days  
Monica L. Correa, SA News Editor • Today, 12:57 PM • 21 Comments  
... **Nvidia's** (NVDA) market cap declined by \$429B in the last three days. In a Goldman Sachs Flow of Funds ...

**BUZZ Investing: U.S. Stocks Rally As Nvidia Climbs Higher**  
VanEck • Today, 6:25 AM  
... of May, Shares of **NVIDIA** Pace Advancing Stocks within the BUZZ Index Shares of **NVIDIA** Corporation (NVDA) ... Additionally, **Nvidia's** announcement of a 10-for-1 stock split, aimed to improve liquidity and make **Nvidia's** shares ... and CPI reports, boosting investor confidence. **Nvidia** led the gains, soaring over 41% thanks to strong ...

Wall Street Breakfast: Cooling Electrified

Hard to understand the contents at first glance

Difficult to understand which article is reliable

\*Logos and screenshots are sourced from the official websites of [Yahoo Finance](#) and [Seeking Alpha](#)

## Quantify the Issue

553

survey participants

- 
- What do you invest in?
  - Are you satisfied with the process?
  - What will improve your decision?  
etc.

## Understand Pains & Needs

17

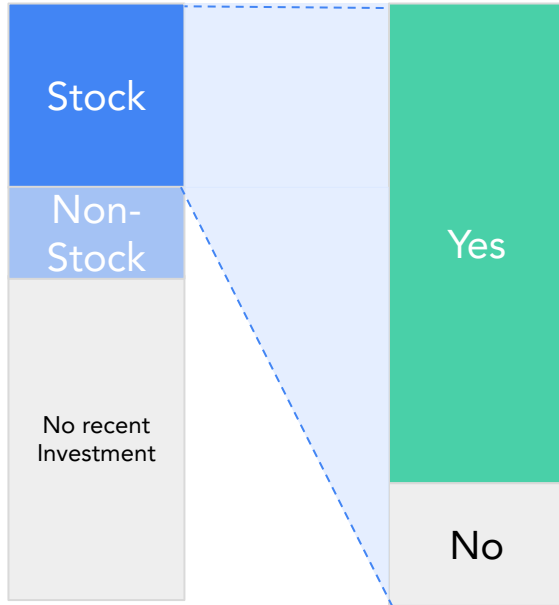
interviewees

- 
- How do you make a decision?
  - How do you use resources?
  - How long do you take for a decision?  
etc.

# Impact Assessment

What did you invest in recently?

Does access to better information sources improve your strategy?



85%

of stock investors want to improve their strategies with more relevant and streamlined information sources

Based on an online survey conducted with 553 participants from July 3 to July 4, 2024.

Questions: 1. What types of investment products, excluding retirement accounts, have you traded in the last 2 years?

2. To what extent do you believe your investment strategy improves with better information sources?

(Responses were on a 5-point Likert scale, with "Strongly Agree" or "Agree" regarded as "Yes.")



02

## Solution Space

As a **retail investor**, I want to make intelligent investment decisions quickly & effortlessly.

For this, I need access to current, comprehensive, precise, and evidence-based intelligence.

## Completeness

One-stop shop for all needed investment process intelligence

## Efficiency

Hit me only with the most important intelligence first. And summarize it for me.

## Evidence

Make sure this intelligence is generated from legitimate sources. Give me the ability to cross-reference from sources.

## DIVITIAE Terminal

### Key Metrics

A numeric snapshot of a stock's basic financial health, including its market value, profitability, and efficiency ratios

### Valuation

A visual representation of a comparison of a company's stock price and net present value, which reflects its intrinsic value

### Performance

A visual representation of a company's revenue and income, which provides insights into its performance and growth potential

### Market hype

A visual representation of the frequency of mentions of a company in news outlets

### News Summary

AI-generated summaries of news articles covering various aspects of a company, with a comprehensive headline

### Earnings Call Summary

AI-generated summaries of a recent earnings call, that helps a holistic performance evaluation of a company

# Demo

---

## Testimonials

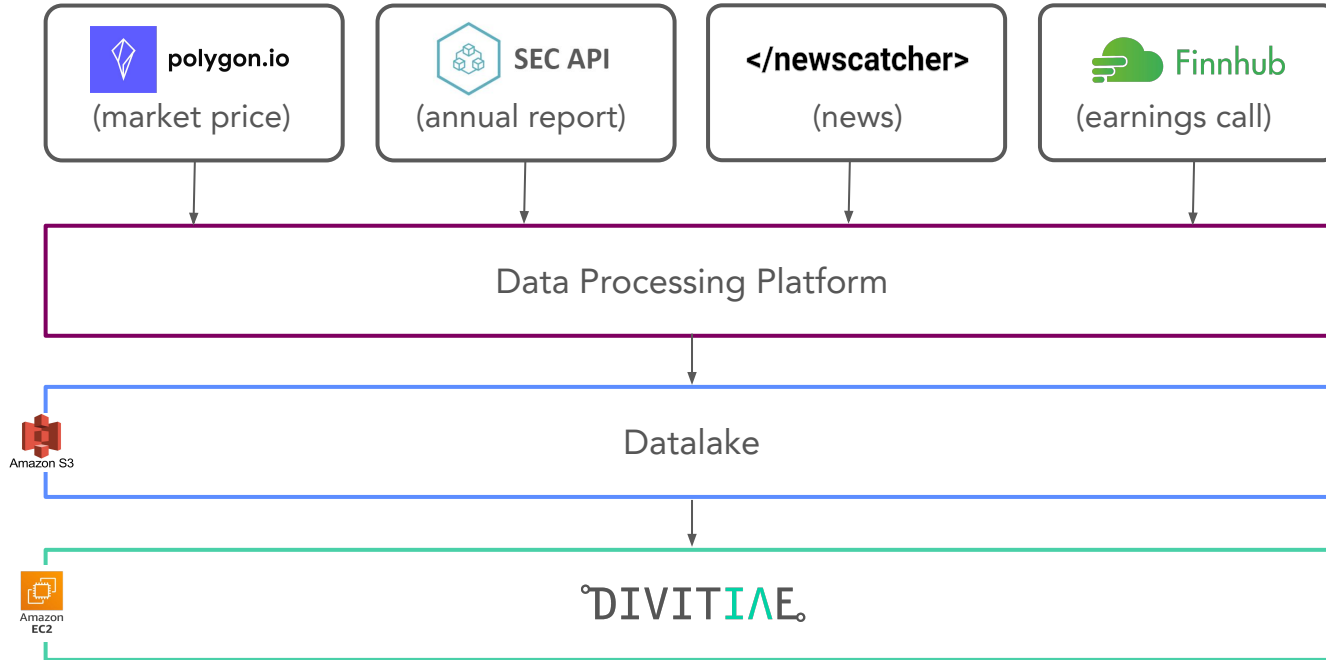
“Loved the simplicity. Does not have to do any math on ratios or calculations. Really impressed by AI-generated summaries.”

“The news and their earnings calls summarization feature isn't available on any other platform.”

# 03

## Overall Architecture

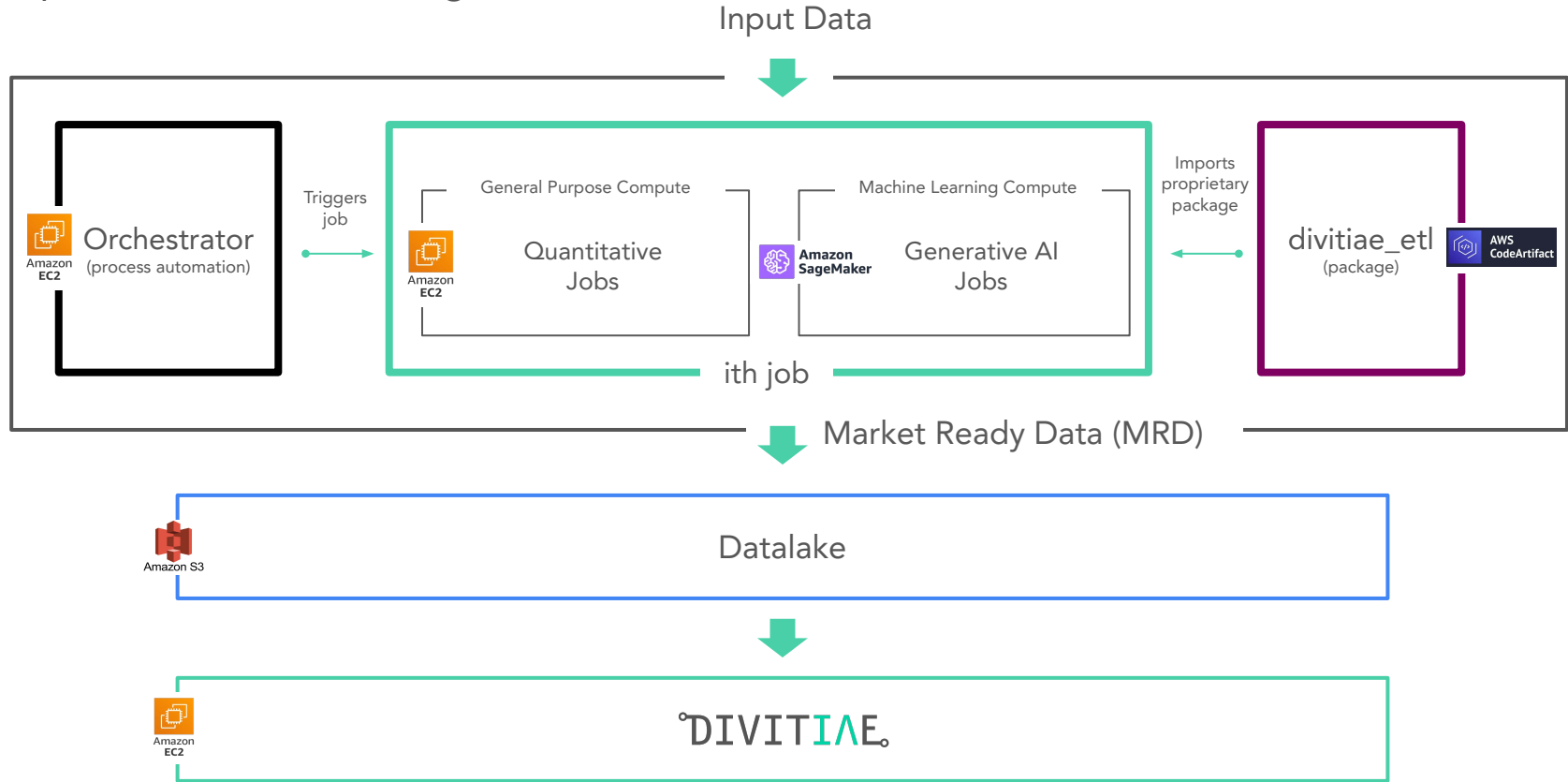
# Technology Architecture Overview



\*Logos are sourced from the official websites of [NewsCatcher, Inc.](#), [Finnhub](#), [Polygon.io, Inc.](#), [SEC API](#), and [Amazon Web Services, Inc.](#)



# Deep Dive: Data Processing Platform





\*Logos are sourced from the official websites of [Amazon Web Services, Inc.](https://aws.amazon.com/)

# 04

## Technology: Quantitative Processing

## Quantitative Data

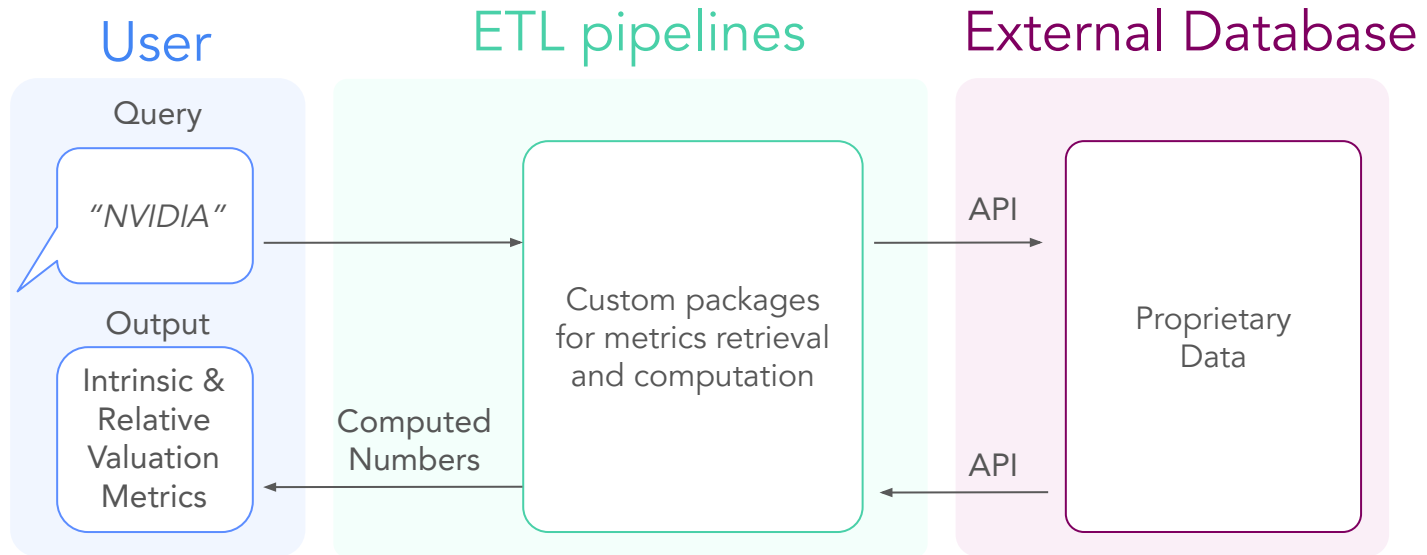
- Developed 12 distinct quantitative models from 2 APIs

Data / API	What's Inside	Usage	EDA Takeaway
 <b>polygon.io</b>	Real-time and historical market data for stocks	Extract stock market prices	<ul style="list-style-type: none"> <li>Real-time market data from all US stock exchanges; returns data frames and custom objects</li> <li>Key features: share prices, volume of shares outstanding for company stock</li> </ul>
 <b>SEC API</b>	Financial and regulatory data for publicly traded companies	Access financial performance from annual / quarterly reports	<ul style="list-style-type: none"> <li>Search and filter company financial statements filed with the SEC</li> <li>Key objects: income statements, balance sheets</li> <li>Key features: revenue, operating income, interest &amp; tax expenses, cash holdings, long-term debt</li> </ul>

\*Logos are sourced from the official websites of [Polygon.io, Inc.](https://polygon.io) and [SEC API](https://sec.gov).

# Understanding the Quantitative Pipeline



- Based on customer queries, the pipeline access external database via API to compute metrics
- Evidence-based Discounted Cash Flow (DCF) assumptions using reported facts
- Complement DCF-based valuation with relative valuation metrics



05

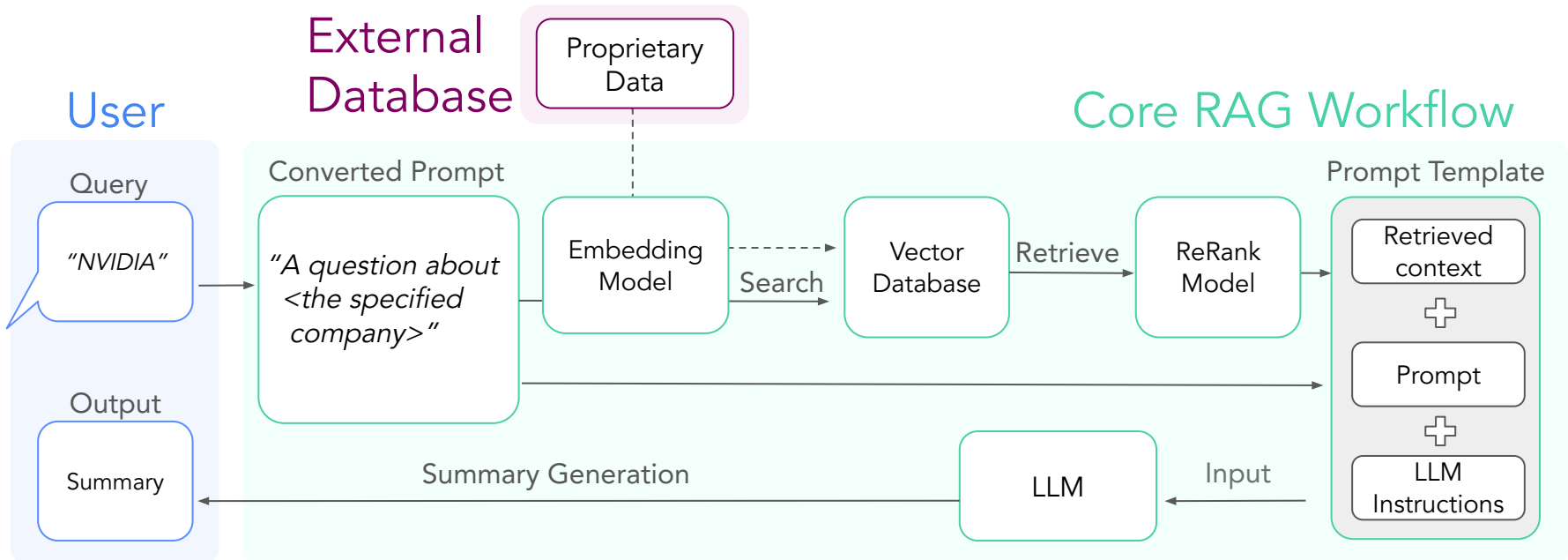
## Technology: Generative AI

# Natural Language Data for RAG

Data / API	What's Inside	EDA Takeaway
	<p>News articles from diverse sources across the web</p>	<ul style="list-style-type: none"> <li>• Dataset is huge. +1M articles / day comes from 70,000 sources in 195 countries and 55 languages.</li> <li>• Clean structure overall, with minor missing values in features like author.</li> <li>• Avg. number of articles differs widely by company . (e.g., Apple: &gt;1000/day, Unilever: &lt;100/day)</li> <li>• Numerous duplicates present, requiring resolution in the pipeline.</li> </ul>
	<p>Financial and regulatory reporting for publicly traded companies</p>	<ul style="list-style-type: none"> <li>• Dataset Contains 15+ years of transcripts for over 65,000 global companies.</li> <li>• Careful engineering was needed as transcripts for other types of events were included alongside earnings calls</li> <li>• Addressed anomalies due to discrepancies between calendar and fiscal years.</li> <li>• Original data was segmented by speakers; concatenation was performed for continuity.</li> </ul>

# RAG Pipeline

- Retrieval Augmented Generation (RAG) is the process of enhancing LLMs by incorporating additional information from external knowledge sources.
- This framework enables us to gather the latest news and earnings call information, which we use to create concise summaries.



Considering the diverse elements of the RAG pipeline and potential future developments, we've adopted a holistic evaluation approach for various modeling combinations.

44

total modeling combinations

- Prompt
- LLM
- Embedding model
- Doc chunk size
- Doc overlap size
- Retrieval quantity
- Rerank quantity

x

47

evaluation metrics

- SBERT Similarity
- SimCSE Similarity
- 
- RAGAS Context Precision
- RAGAS Answer Similarity
- RAGAS Answer Correctness







# Best RAG Component Combination

## Best Modeling Combination

Component / Parameter	Solution / Value
Prompt	v2
LLM	ChatCohere
Embedding Model	GISTEmbed
Doc Chunk Size	512
Doc overlap size	128 (1/4 of the total)
Retrieval quantity	200
Rerank quantity	50



## Score of the Focus Evaluation Metrics

Evaluation Metric	Objective	Score	
		News	Earnings Calls
SBERT Similarity	Semantic similarity	0.85	0.91
SimCSE Similarity	Sentence-level similarity	0.86	0.91
RAGAS Context Precision	Accuracy of context inclusion	0.86	0.99
RAGAS Context Recall	Completeness of contextual information	0.61	0.72
RAGAS Answer Relevancy	Relevance of summary	0.60	0.67
RAGAS Faithfulness	Compares similarity	0.81	0.95
RAGAS Answer Similarity	Similarity	0.88	0.94
RAGAS Answer Correctness	Factual correctness	0.67	0.65

06

## Looking Forward

# Key Challenges

Design Area	Challenge	Solution
Technology Architecture	Integrating various technologies	Automation of programmatically generated ML environments
AI Engineering	Mastering new RAG and LLM technologies	Tested 44 model combinations against 47 key evaluation metrics
Problem Space & Customer Development	Narrowing problem space & target customers	Continue customer development (~600 interviews & survey responses)
Intellectual Property Law	Managing risk given evolution of legal landscape for AI generated content.	Including data summary sources, monitoring legal decisions.

# Product Roadmap

Design Area	User Story	Implementation
Scale & Coverage of Intelligence	As a user, I want to be able to see more intelligence per company, as well as all companies in a stock exchange	Generalize solution from subset of publicly traded companies for all publicly traded
Cohort Analysis	As a user, I want to compare, cluster, and rank different companies by category	Develop and deploy comparison metrics and visualization components to the user interface
User Education	As a user, I need help understanding financial terms and visuals related to investment research	Add an education component that fosters financial literacy
Accessibility	As a user with disabilities, I want to enjoy the value propositions of this product	Ensure multi-platform compatibility to increase accessibility

## Our Mission

Democratizing financial information,  
by empowering all with a wealth of intelligence  
and intelligence of wealth

## Solution

DIVITIAE.AI simplifies investment research with key features, saving time and increasing investment confidence



## Acknowledgements

We would like to express **our gratitude** to:

- NewsCatcher Inc. and Finnhub for generously providing us with free access to their data.
- Interview Participants for sharing valuable insights and suggestions with us.
- Instructors Joyce Shen and Zona Kostic for their consistent and instrumental guidance throughout the project.
- NLP instructors Mark Butler and Natalie Ahn for their extensive hands-on advice on NLP development.
- All Capstone classmates for their constructive feedback and inspiring ideas.

# Appendix

---

# Details of Discounted Cash Flow & Relative Valuation

## INTRINSIC VALUATION

### Discounted Cash Flow

$$DCF = \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \frac{CF_n}{(1+r)^n}$$

where:

$CF_1$  = The cash flow for year one

$CF_2$  = The cash flow for year two

$CF_n$  = The cash flow for additional years

$r$  = The discount rate

### The discount rate: Weighted Avg Cost of Capital

**WACC** = (%Equity × Cost of Equity) + (%Debt × After-Tax Cost of Debt)

where:

**%Equity** = Value of Firm's Equity / Total Value of Capital (Equity + Debt)

**%Debt** = Value of Firm's Debt / Total Value of Capital (Equity + Debt)

**Cost of Equity** =  $R_f + \beta \times R_m - R_f$

where  $R_f$  = Risk-Free Rate,

$R_m$  = Market Return

$\beta$  = Covariance(Company Stock Return,  $R_m$ ) ÷ Variance( $R_m$ )

**Cost of Debt** = Interest ÷ Debt × (1 - Tax Rate)

**To forecast future cash flows:** use historical growth rate of net income

**To compute discount rate:** use the following assumptions

- **Risk-free rate pinned to 4%** (approximate yield on U.S. 10 Year Treasury Bond)
- **Market return pinned to \$SPY** (index fund that approximates S&P 500)
- $\beta$  (**beta**) is a measure of a stock's volatility vs all other stocks in the market

## RELATIVE VALUATION

**Market Cap:** Last Share Price × # of Shares Outstanding

**Enterprise Value (EV):** Market Cap + Long-Term Debt - Cash ✓ Cross-company valuation compares

**Earnings Per Share:** Operating Income ÷ # of Shares Outstanding ✓ Complements DCF-based valuation

**Price-to-Earnings (P/E):** Last Share Price ÷ Earnings Per Share