



# FROM HYPE TO PRODUCTION

THE BLUEPRINT TO DELIVER GEN AI USE CASES IN FINANCIAL SERVICES

**JONATHAN REGENSTEIN**

# Safe Harbor and Disclaimers

Other than statements of historical fact, all information contained in these materials and any accompanying oral commentary (collectively, the “Materials”), including statements regarding (i) Snowflake’s business strategy and plans, (ii) Snowflake’s new or enhanced products, services, and technology offerings, including those that are under development or not generally available, (iii) market growth, trends, and competitive considerations, and (iv) the integration, interoperability, and availability of Snowflake’s products, services, or technology offerings with or on third-party platforms or products, are forward-looking statements. These forward-looking statements are subject to a number of risks, uncertainties and assumptions, including those described under the heading “Risk Factors” and elsewhere in the Annual Reports on Form 10-K and the Quarterly Reports on Form 10-Q that Snowflake files with the Securities and Exchange Commission. In light of these risks, uncertainties, and assumptions, the future events and trends discussed in the Materials may not occur, and actual results could differ materially and adversely from those anticipated or implied in the forward-looking statements. As a result, you should not rely on any forwarding-looking statements as predictions of future events.

Any future product or roadmap information (collectively, the “Roadmap”) is intended to outline general product direction; is not a commitment, promise, or legal obligation for Snowflake to deliver any future products, features, or functionality; and is not intended to be, and shall not be deemed to be, incorporated into any contract. The actual timing of any product, feature, or functionality that is ultimately made available may be different from what is presented in the Roadmap. The Roadmap information should not be used when making a purchasing decision. In case of conflict between the information contained in the Materials and official Snowflake documentation, official Snowflake documentation should take precedence over these Materials. Further, note that Snowflake has made no determination as to whether separate fees will be charged for any future products, features, and/or functionality which may ultimately be made available. Snowflake may, in its own discretion, choose to charge separate fees for the delivery of any future products, features, and/or functionality which are ultimately made available.

© 2024 Snowflake Inc. All Rights Reserved. Snowflake, the Snowflake logo, and all other Snowflake product, feature and service names mentioned in the Materials are registered trademarks or trademarks of Snowflake Inc. in the United States and other countries. All other brand names or logos mentioned or used in the Materials are for identification purposes only and may be the trademarks of their respective holder(s). Snowflake may not be associated with, or be sponsored or endorsed by, any such holder(s).





## Agenda

- Key Requirements for a Gen AI Strategy
- Financial Services Use Cases
- Building on Snowflake
- Q&A

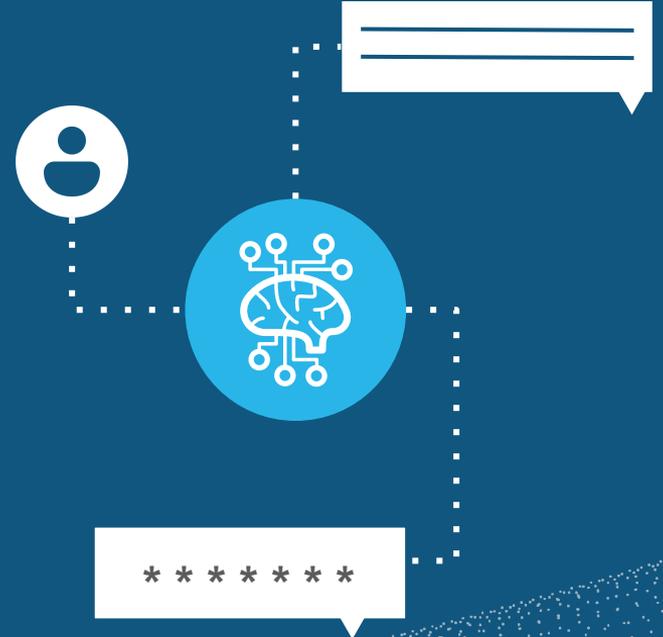


**JONATHAN  
REGENSTEIN**

Head of AI / ML, Financial  
Services

# We are in the early stages of a new wave of AI assistants accessible to everyone in natural language

Summarization, generation, and translation assistants using Large Language Models (LLMs) can help you reimagine every aspect of your business.



# BUSINESS CHALLENGES STEM FROM TECH ONES

## BUSINESS CHALLENGES

### Trust & Compliance

Data quality, privacy, and security

### Total Cost of Ownership (TCO)

Many projects, new tools and infrastructure, few skilled resources

### Proliferation of shadow IT

Growth in no-code apps using LLMs for business users

## TECHNOLOGY CHALLENGES

Data silos &  
Model governance

Infrastructure ops &  
integration maintenance

Internal & 3rd-party  
AI app governance

# KEY REQUIREMENTS



## Scale of Data

Access to volumes of data is required to train or fine-tune large language models



## Data Quality

Poor data quality can lead to data bias. As with any other computer system, AI systems are only as good as the data



## Compute Power

Enormous processing power needed to build AI systems and leverage ML, image processing or language understanding



## Security & Governance

Strict security controls are needed for both model training and model use, with focus on intellectual property, PII, data exfiltration



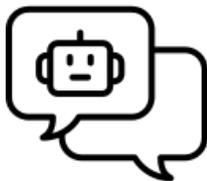
## Technical Expertise

Training LLMs require an understanding of deep learning workflows, transformers, distributed software and hardware



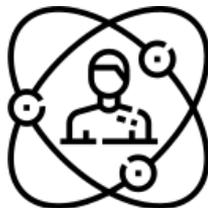
# Financial Services Use Cases

# POPULAR GENERATIVE AI USE CASES



## Conversational Co-Pilot for Data Analysis

Improve employee and customer productivity with AI assistants to help identify trends in your data.



## Accelerating Employee Expertise

Scale learning and development efforts with AI-powered tools trained with your business knowledge.



## Document Analysis and Summarization

Rapidly gain insights from large volumes of unstructured data without manual human labor.

# POWERING AUGMENTATION IN BANKING

According to recent research, the banking sector represents the top industry (across all verticals) for generative AI to automate or augment employee's work.

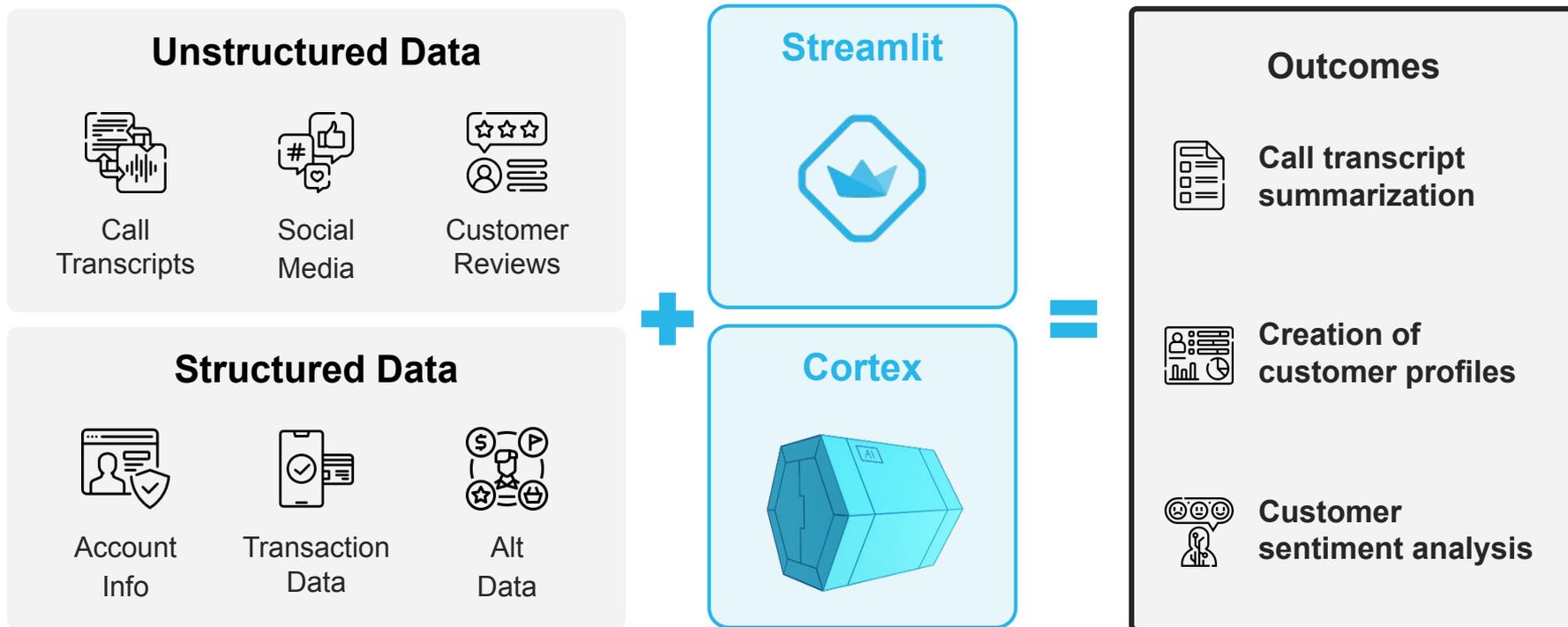


# 37%

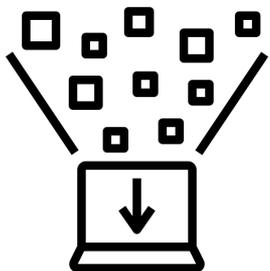
of customer service rep time in banking right now could be augmented by generative AI



# Banking Use Case: Call Center Optimization



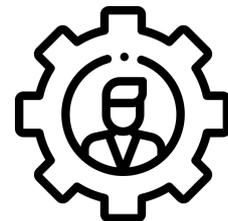
# WEALTH & ASSET MANAGEMENT TRENDS



**ACCESS TO MORE  
ALTERNATIVE &  
UNSTRUCTURED DATA**



**FASTER RESEARCH &  
ALPHA GENERATION**



**EVOLVING INVESTOR  
EXPECTATIONS**

# USE CASE: QUANTITATIVE RESEARCH

## Ability to scan through thousands of research

J.P.Morgan

North America Equity Research  
30 September 2015

### 21st Century Fox

Hulu Steadily Building a Streaming 'Empire' as New Tier Debuts and More Hits Roll On

**Overweight**  
FOXA, FOXA US  
Price: \$25.80  
Price Target: \$33.00

**Media**  
**Alexis S. Quadracci, CFA**  
(1-212) 852-1588  
www.aquadracci@jpmorgan.com  
Bloomberg: JPMO\_AQUADRACI <AQ>  
**James Koppelman**  
(1-212) 219-5880  
james.koppelman@jpmorgan.com  
**David Karnovsky, CFA**  
(1-212) 602-1058  
david.karnovsky@jpmorgan.com  
**Julia Yee**  
(1-212) 602-8986  
julia.yee@jpmorgan.com  
J.P. Morgan Securities LLC

**Price Performance**

**Price Performance**  
2012 2013 2014 2015  
25% 20% 15% 10% 5% 0% -5% -10% -15% -20%

**Twenty-First Century Fox, Inc. (FOXA/FOXA US)**

FYE Ann	2013A	2014A	2015A	2016E	Company Data	
EPS - Reporting (\$)	0.38	0.33	0.39	0.39	Price (\$)	25.80
Q1 (Sep)	0.10	0.13	0.41	0.42	Date of Price	29 Sep 15
Q2 (Dec)	0.32	0.47	0.42	0.46	52-week Range (\$)	39.27-22.81
Q3 (Mar)	0.31	0.43	0.29	0.45	Market Cap (\$ m)	53,149.00
Q4 (Jun)	0.25	0.29	0.45	0.45	Float (m)	2,260
P/E	1.46	1.59	1.66	1.55	Share Cto (m)	2,260
CY	1.46	1.59	1.66	1.55	Price Target (\$)	33.00
Dividends (EPS FY (\$))	0.00	0.00	0.00	0.00	Price Target End Date	31-Dec-15
Source: Company data, Bloomberg, J.P. Morgan estimates.	1.53	1.70	1.82			



## Access to other structured & semi-structured data

-  Pricing Data
-  Reference Data
-  Fundamental Data
-  Private Asset Data
-  Alternative Data
-  Economic Data
-  ESG Data
-  Quant Analytics



## Outcomes



Productivity gains with more efficient quant workflows



Faster ability to synthesize lots and lots of data



Cost efficiencies with research automation



# INSURANCE: UNSTRUCTURED DATA

## Personal Lines Carriers



## Commercial Lines Carriers



## Life & Annuities



## Specialty Carriers



## Agents / Brokers



## Reinsurers



Policy Documents



Claim Notes



Medical Notes



Contracts



Aerial Imagery



Audio Files



Emails



# USE CASE: DOCUMENT AI

Document AI simultaneously comprehends all aspects of documents:

## SEMANTICS

Document type: Application for LT Disability Income Benefits

## CONTEXT

Customer: Sully McConnell  
Company Name: Snowflake

## LAYOUT

Basic Amount: 20,000.00

## GRAPHICS USING MULTI-MODAL LLM

Customer Initials: Yes  
Seller's Initials: Yes

Fax or mail the completed application to:  
The Hartford  
P.O. Box 14301 HARTFORD LIFE AND ACCIDENT INSURANCE COMPANY  
Lexington, KY, 40512-4301  
Fax Number: (877) 431-8901

**APPLICATION FOR LONG TERM DISABILITY INCOME BENEFITS**

**Section I - Employer's Section - To be Completed by the Employer**

This claim is for (Employee's Name): Sully McConnell Social Security Number: 123-123-123 Date of Birth: 02/14/2024

Employee's Address: (Street, City, State, Zip)  
114 W 41st St, New York, NY 10036

**A. Information About the Employer**  
Company's Name: Snowflake Group Policy Number:  
Address: (Street, City, State, Zip) Telephone Number: (123) 123-4567 Fax Number: (123) 123-4567  
114 W 41st St, New York, NY 10036  
Name and address of division where employee works: (if different from above) Class: Location:

**B. Information About the Employee**  
Date employee was hired: 02/14/2024 Date employee became insured under this plan: 02/14/2024 What was the employee's regularly scheduled work week? 60 hours per week.  
Was the employee's LTD insurance issued on the basis of a Personal Health Statement?  Yes  No If "Yes," attach copy.  
Was the employee insured under your prior LTD policy?  Yes  No If "Yes," please provide the inclusive date of coverage. From 02/14/2023 Through 02/14/2024 Has the employee been terminated?  Yes  No If "Yes," date.  
Reason: N/A  
Was the employee on Qualified Family Leave when disability began?  Yes  No  
Did LTD insurance continue while on Family Leave?  Yes  No  
Date Leave of Absence started under Family Leave Act:

**C. Information for Group Life Insurance Waiver Benefits**  
Does the employee also have Group Life Insurance coverage with The Hartford?  Yes  No If "Yes," provide the following information: Basic Amount \$ 20,000 Supplemental Amount \$  
Effective Date of Group Life Insurance coverage: 02/14/2024

**D. Information Needed for Withholding and Reporting Taxes**  
What percent of this employee's LTD benefits is taxable? 10%  
What percentage, if any, do you contribute towards the cost of the LTD premium? 20%  
Does the employee contribute towards the cost of the LTD premium?  Yes  No.  
If "Yes," is it on a  Pre or  Post Tax basis?

**E. Information About the Claim**  
Were there any changes to the employee's job responsibilities due to the disabling condition before the employee became totally disabled?  Yes  No If "Yes," what were the changes, and when were they made?  
Sully now has to present on stage at SKO 2024  
What was the employee's permanent job on his or her last day at work?  
N/A How long has the employee been in this job?  
2  
Why did employee stop working? Stage fright from having to present at SKO Is the employee's condition work related?  
 Yes  No  
Last day employee actually worked: On that day, did the employee work a full day?  Yes  No  
If "No," how many hours were worked?  
Has a claim been filed with Workers' Compensation?  Yes  No Date employee is expected/did return to work:  
If "Yes," send initial report of illness or injury and award notice. Full time?  Yes  No  
Name and address of your compensation carrier

**F. Information About Your Pension Plan (Do not complete for maternity claim.)**  
Do you have a pension plan?  Yes  No If "Yes," what type? (Check as many as applicable)  
 Defined contribution  Profit Sharing  Defined benefit  401 K  Other (specify)  
Is the employee eligible for your pension plan?  Yes  No If eligible, does the employee participate?  Yes  No  
If "No," why?  
If the employee is participating, when is he or she eligible for benefits under the plan? Yes  
At what point does the employee qualify for a full pension?  
Is there a Disability Retirement Option available to this employee?  Yes  No

LC-4571-34 Page 2 of 10 01/2013

# Business Objectives Across Insurance

## ORGANIZATIONAL COST REDUCTION AND EFFICIENCIES

## CONSTITUENCY EXPERIENCE VIA CO-PILOT



Claims



Underwriters



Customer Service  
Reps (CSRs)



Internal Agent /  
Broker Co-pilot



Internal CSR  
Co-pilot

# Building on Snowflake



# FLEXIBILITY & OPTIONALITY

From cloud providers to data and service providers, and now to LLMs providers, Snowflake is extending our strategy of offering optionality and best-of-breed capabilities for financial services.



# SNOWFLAKE PLATFORM OVERVIEW

Access

Build

Deploy

LLM access: Seamless, Open and Flexible

**Snowflake Cortex:**  
Access **fully hosted** LLMs



**Snowpark Container Services:** Build or share proprietary LLMs



**Snowflake Marketplace:**  
Access Partner LLMs

**AI21 labs**

**External Functions Call**  
to leverage external LLMs



**Snowflake Copilot:**  
Use Gen AI to write code

App Layer



**Streamlit:**  
Build a UI for your Gen AI chat app

Vector Embedding & Cortex Search (i.e. the Orchestration Layer)

Data access to structured, semi-structured and unstructured data

**Document AI:**  
Tap into unstructured data



**Snowflake Marketplace:**  
Access 3rd party structured data

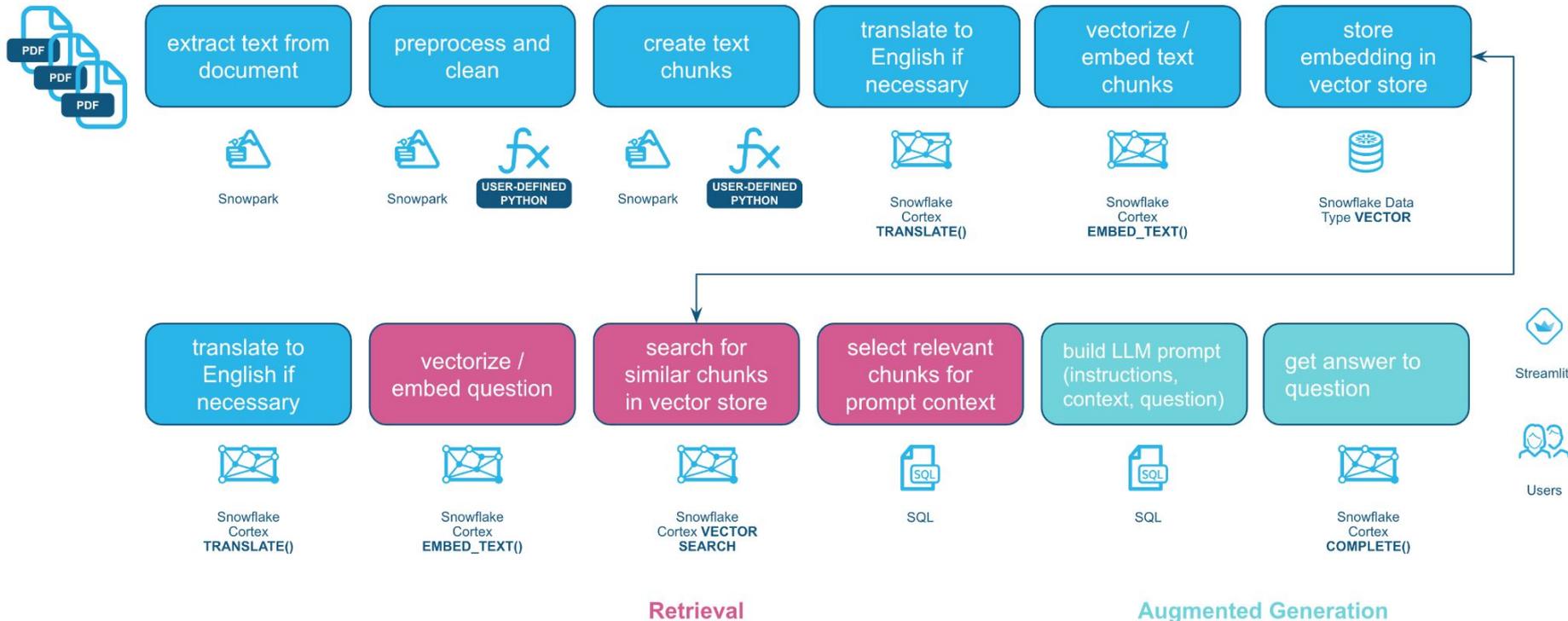


**Iceberg Tables:**  
Access data across the enterprise



Snowflake-Governed Data

# Retrieval Augmented Generation in Snowflake





# LLMs and GenAI

## (recent and upcoming improvements)

# Cortex Fine-Tuning

## What Is It

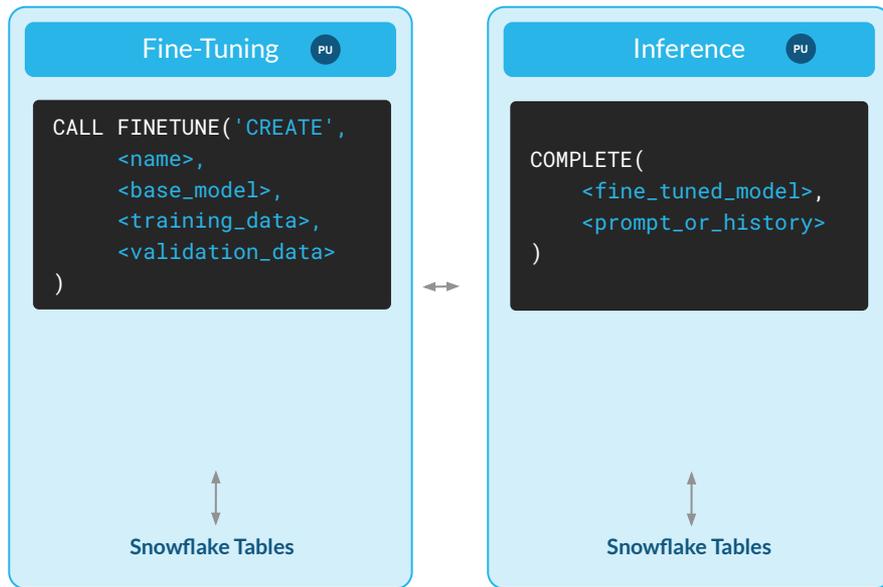
Serverless SQL/Python functions that fine-tune and run inference on LLMs

## Why Use It

Receive specific outputs from unstructured text data in a cost-effective way using smaller, task-tuned language models

## How To Use It

Training: Use fine-tuning API (PrPr) or UI (PuPr)  
Inference: Use SQL/Python function in Snowsight or expose as a custom app via Streamlit



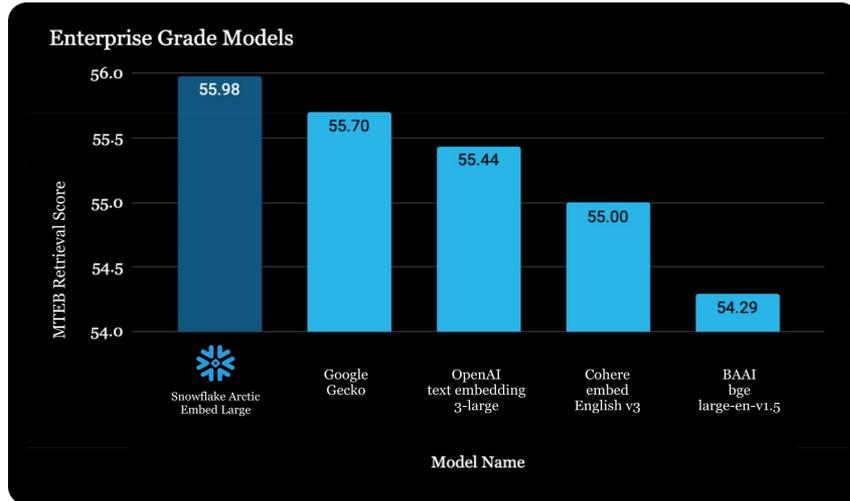
PR Private Preview

PU Public Preview



# Arctic Embedding Models

- Suite of models available in five sizes ranging from x-small (xs) to large (l)
- State-of-the-art retrieval performance on the Massive Text Embedding Benchmark
- Large (l) model outperforms closed-source models estimated to be roughly 4x in size
- Medium (m) model includes long-context support long document retrieval with extended context support of up to 8192 tokens.



# Cortex Search: a performant, hybrid search engine for unstructured data

## What is it?

**Hybrid search engine:** leverages a fusion of vector embeddings for semantic similarity, plus keyword search for lexical similarity. Cutting-edge retrieval capabilities **built on Neeva search**

**Fully managed by Snowflake:** secure and governed within your Snowflake account, exposed via REST, Python APIs

## Creating the service:

```
CREATE CORTEX SEARCH SERVICE mysvc
ON text
ATTRIBUTES date,region
WAREHOUSE = xs_wh
TARGET_LAG = '1 minute'
AS ( SELECT text, date
FROM transcripts );
```

## Querying the service:

```
curl https://mysvc...
-d '{
  "query": "ecb regulations",
  "filter": "region='EMEA'",
  "limit": 10,
}'
```

# Cortex Search unlocks scalable, business-specific LLM apps in Snowflake



## Chat with my documents

Build your own **proprietary chat experiences** for your business users and customers, **grounded in business-specific context** extracted from your documents

**Get more value out of your underutilized text data**, save countless hours on analyst activities



## Search for my documents

Build your own **enterprise search engine for your documents**.

Get up and running with **state-of-the-art retrieval capabilities in minutes**, no tuning required.

**Reduce analyst time spent** searching for the right business artifact

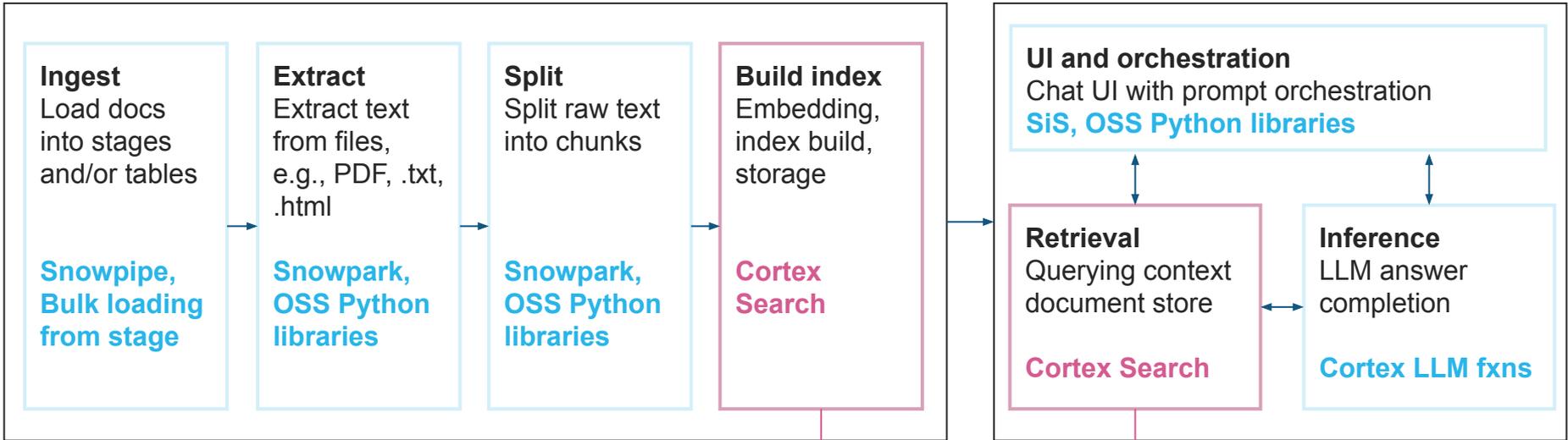


# RAG reference architecture in Snowflake

End-to-end “chat with your document” apps built and served in Snowflake

## Build

## Serve



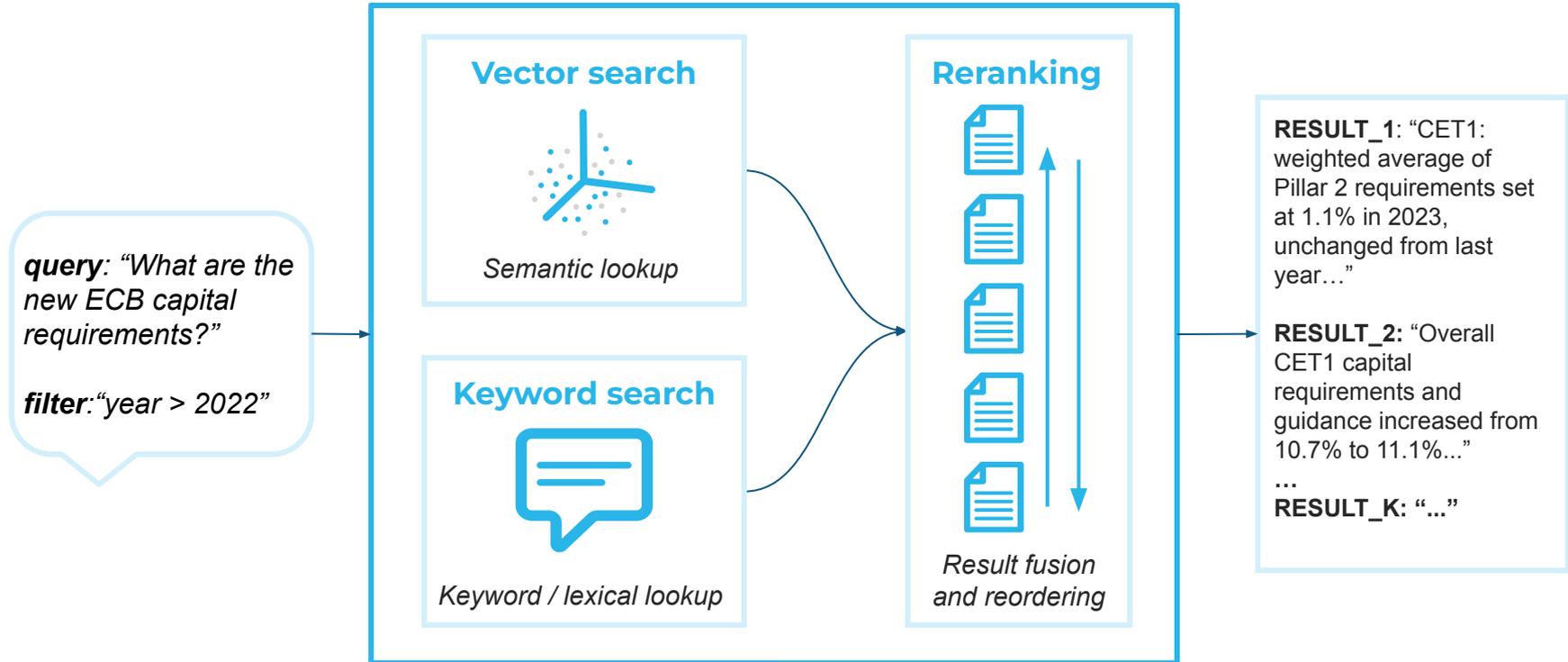
**Cortex Search is the RAG engine for LLM Apps in Snowflake**

# The life of a query in Cortex Search

## User query

## Hybrid Retrieval

## Result set



# THANK YOU

