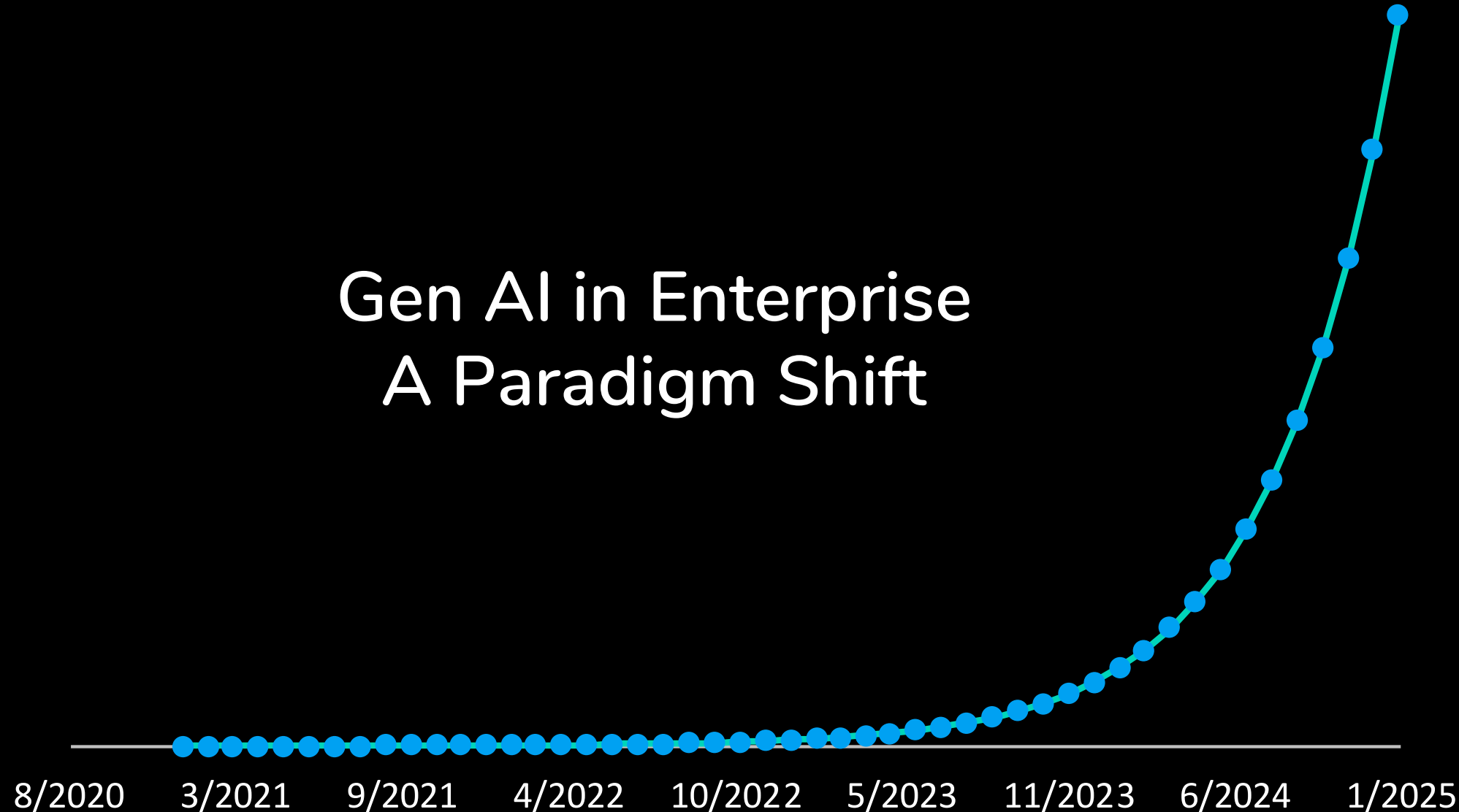


Gen AI in Enterprise A Paradigm Shift



What We'll Cover

1. Trends & Adoption
2. Build vs Buy
3. Open Source
4. Gen AI Design for Enterprise
5. Case Study Examples



1. Trends & Adoption

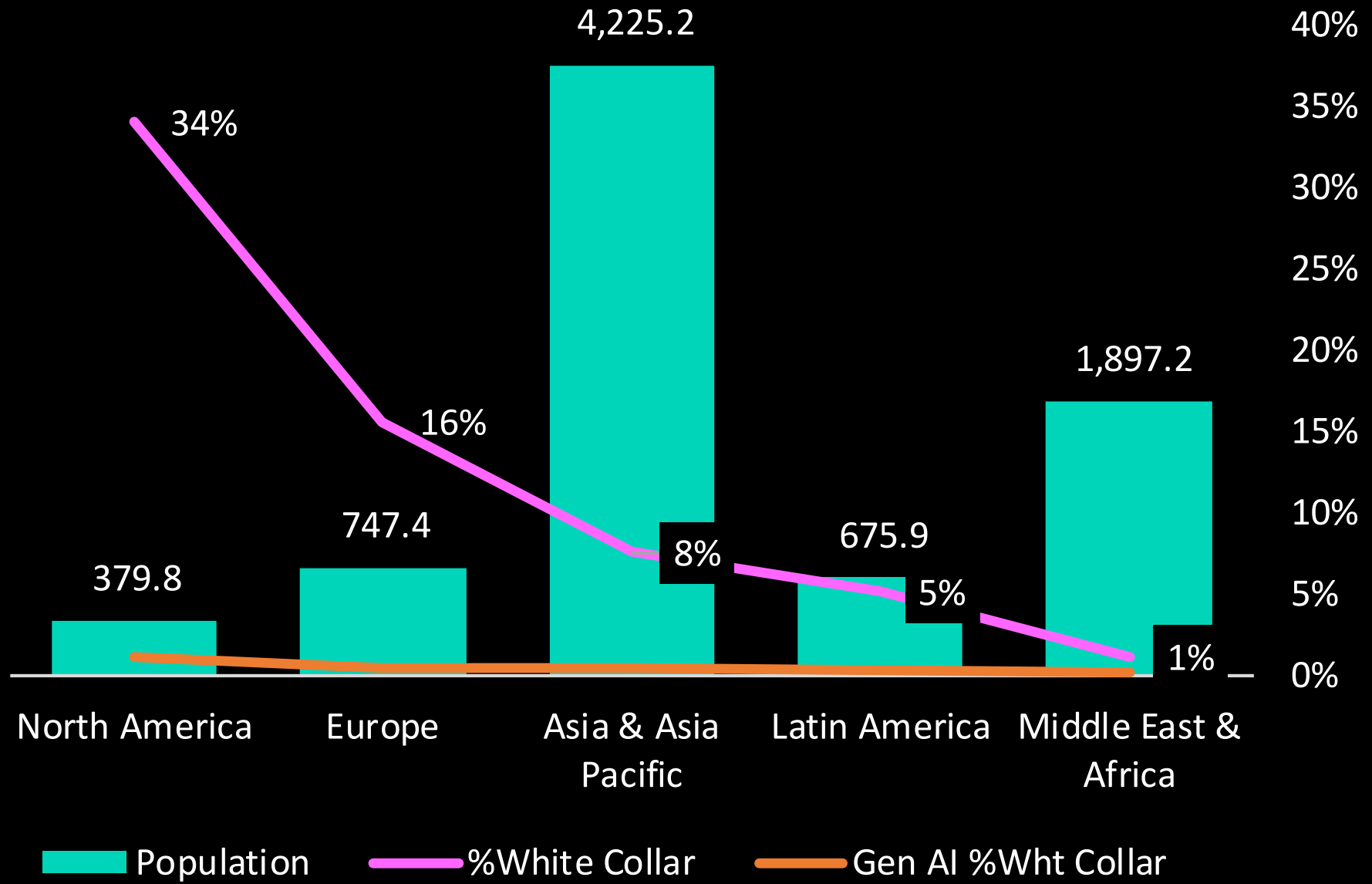
2. Build vs Buy

3. Open Source

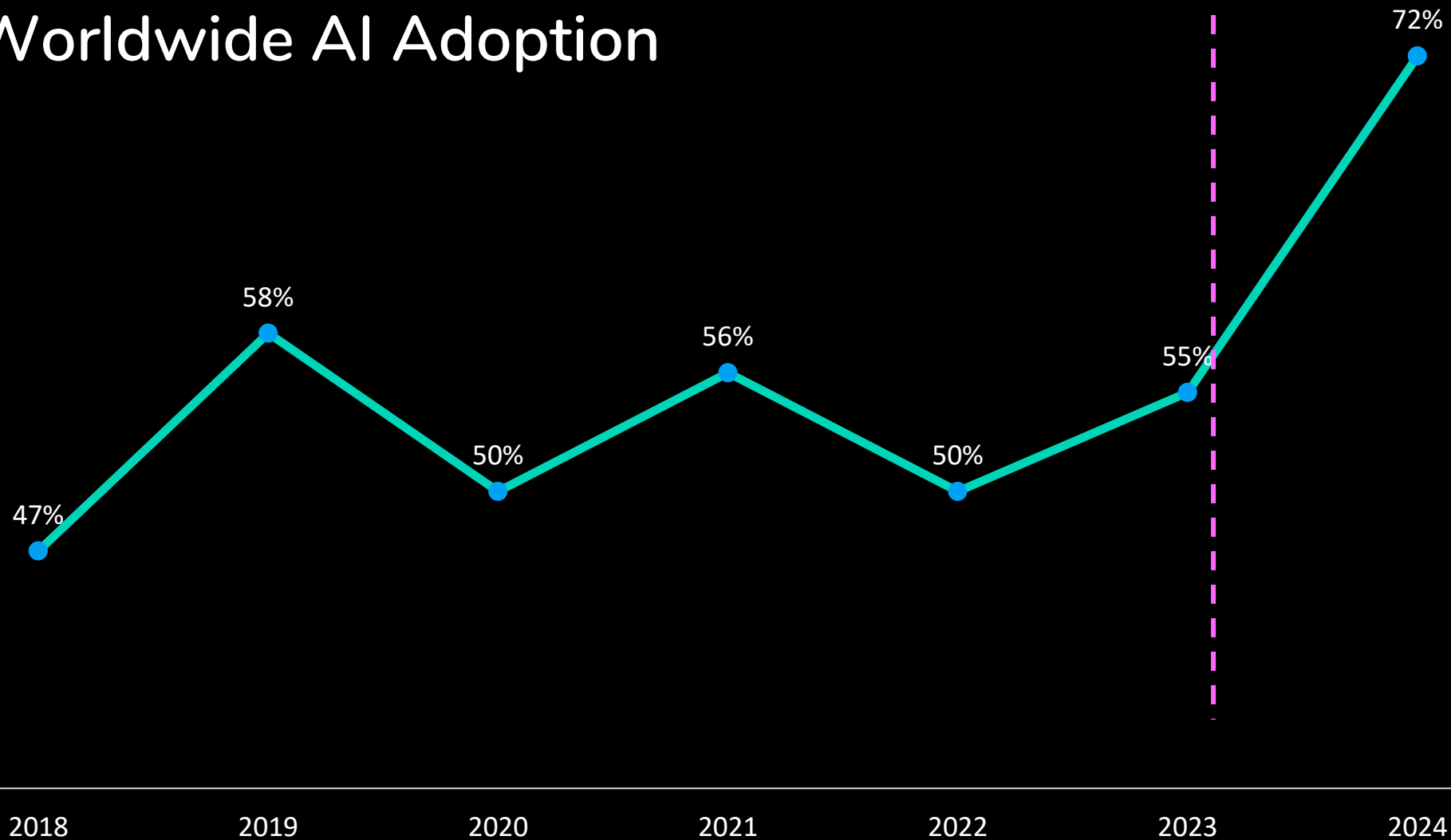
4. Gen AI Design for Enterprise

5. Case Study Examples

Millions			
Region	Population	White Collar	Gen AI Corp Users
North America	379.8	129.3	1.5
Europe	747.4	116.9	0.6
Asia & Asia Pacific	4,225.2	321.2	1.7
Latin America	675.9	35.0	0.1
Middle East & Africa	1,897.2	22.6	0.0
Total	7,925.4	625.1	3.8

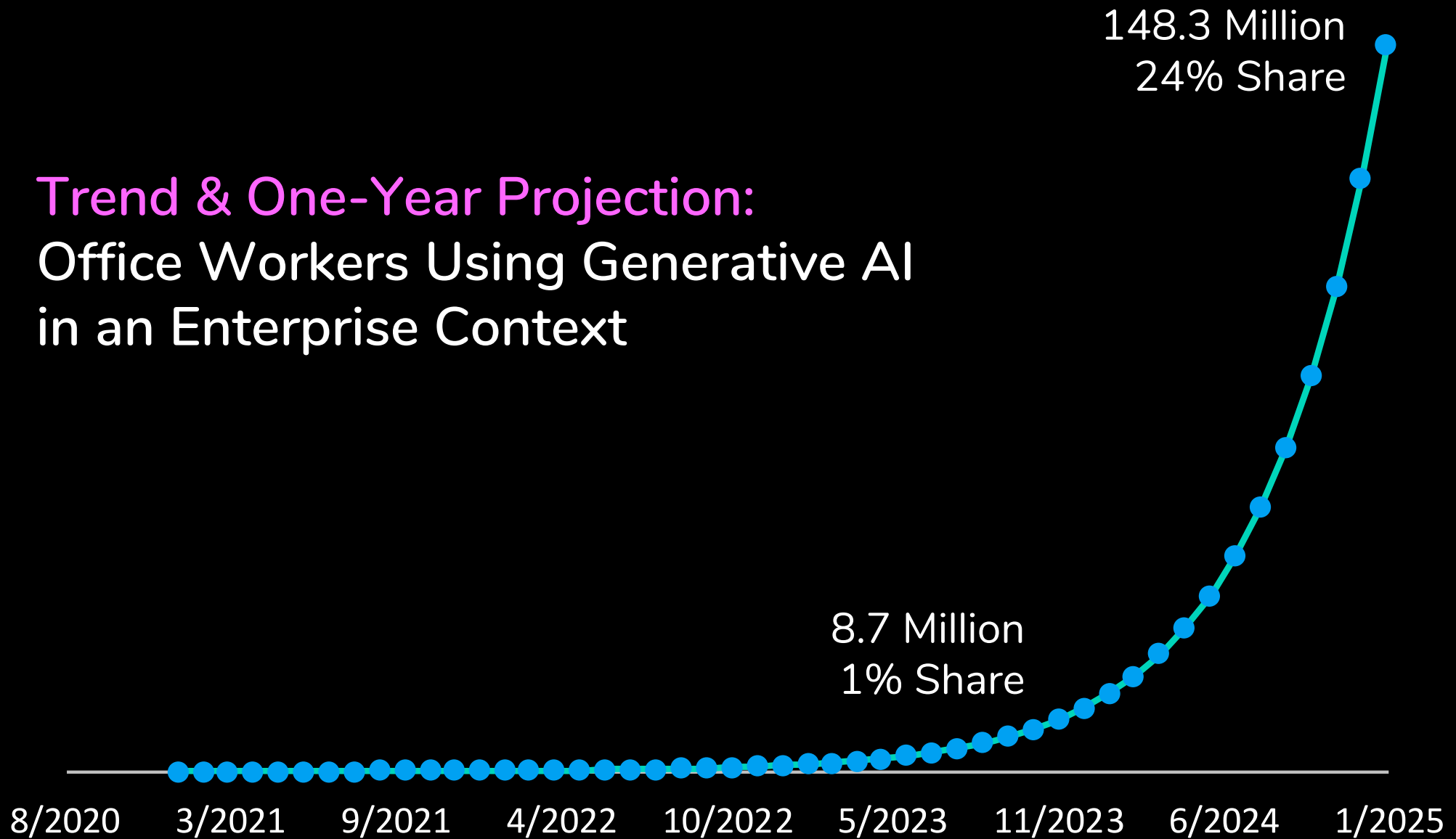


Worldwide AI Adoption



Source: McKinsey, The State of AI (Early 2024)
<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

Trend & One-Year Projection:
Office Workers Using Generative AI
in an Enterprise Context





-
1. Trends & Adoption
 - 2. Build vs Buy**
 3. Open Source
 4. Gen AI Design for Enterprise
 5. Case Study Examples

Home

Shorts

Subscriptions

You >

Your channel

History

Playlists

Your videos

Watch later

Liked videos

Your clips

Subscriptions

OpenAI

Data Phoenix Events

All subscriptions

AI Master Group

@AIMasterGroup · 16.7K subscribers · 53 videos

More about this channel ...more

Home Videos Podcasts Playlists Community

For You

AI MASTER GROUP

Celia Wanderley
Chief Innovation Officer
Bits in Glass

Jim Griffin

30:55

167 views · 1 day ago

⋮

AI MASTER GROUP

Dave Stern
Fractional CTO & DevOps Engineer
Author of Hackproof Your Startup

Jim Griffin

24:13

153 views · 3 days ago

⋮

AI MASTER GROUP

Jodi Blomberg
VP Data Science
Cox Automotive

Jim Griffin

30:57

20 views · 6 days ago

⋮

Videos

AI MASTER GROUP

Celia Wanderley
Chief Innovation Officer
Bits in Glass

Jim Griffin

30:55

AI MASTER GROUP

Dave Stern
Fractional CTO & DevOps Engineer
Author of Hackproof Your Startup

Jim Griffin

24:13

AI MASTER GROUP

Andre Lapinskas, PhD
Director of Search, AI Labs Huawei?
Former VP Engineering and AI, eBay

Jim Griffin

38:01

AI MASTER GROUP

Shawn Goodie
Global VP Software & Partnerships
Facebook

Jim Griffin

46:22

AI MASTER GROUP

Jodi Blomberg
VP Data Science
Cox Automotive

Jim Griffin


30:57

Your Library

Podcasts & Shows

Rechts

AI Master Group
Podcast • Jim Griffin



Podcast

AI Master Group

Jim Griffin

Following


Up next


- Celia Wanderley: AI Innovator of the Year

AI Master Group

Celia is the Chief Innovation Officer at Bits In Glass (BIG), a top Canadian IT consulting firm. Prior to that, she held senior leadership roles at AltaML and at Deloitte Canada. Celia was recognized as the AI Innovator of th...

Oct 12 • 30 min 54 sec





About

Hosted by Jim Griffin, this show features the researchers and leaders who are building the future of AI in their work today. Join us for thought-provoking conversations about the evolving world of artificial intelligence.

Show less

No rating ☆ • Technology

All Episodes

AI MASTER Celia Wanderley: AI Innovator of the Year

AI MASTER

GROUP



Jodi Blomberg
VP Data Science
Cox Automotive

Jim Griffin





Jodi Blomberg
VP Data Science
Cox Automotive

Build if . . .

- Proprietary Data
- Multiple sources
- Revenue Trumps Efficiency

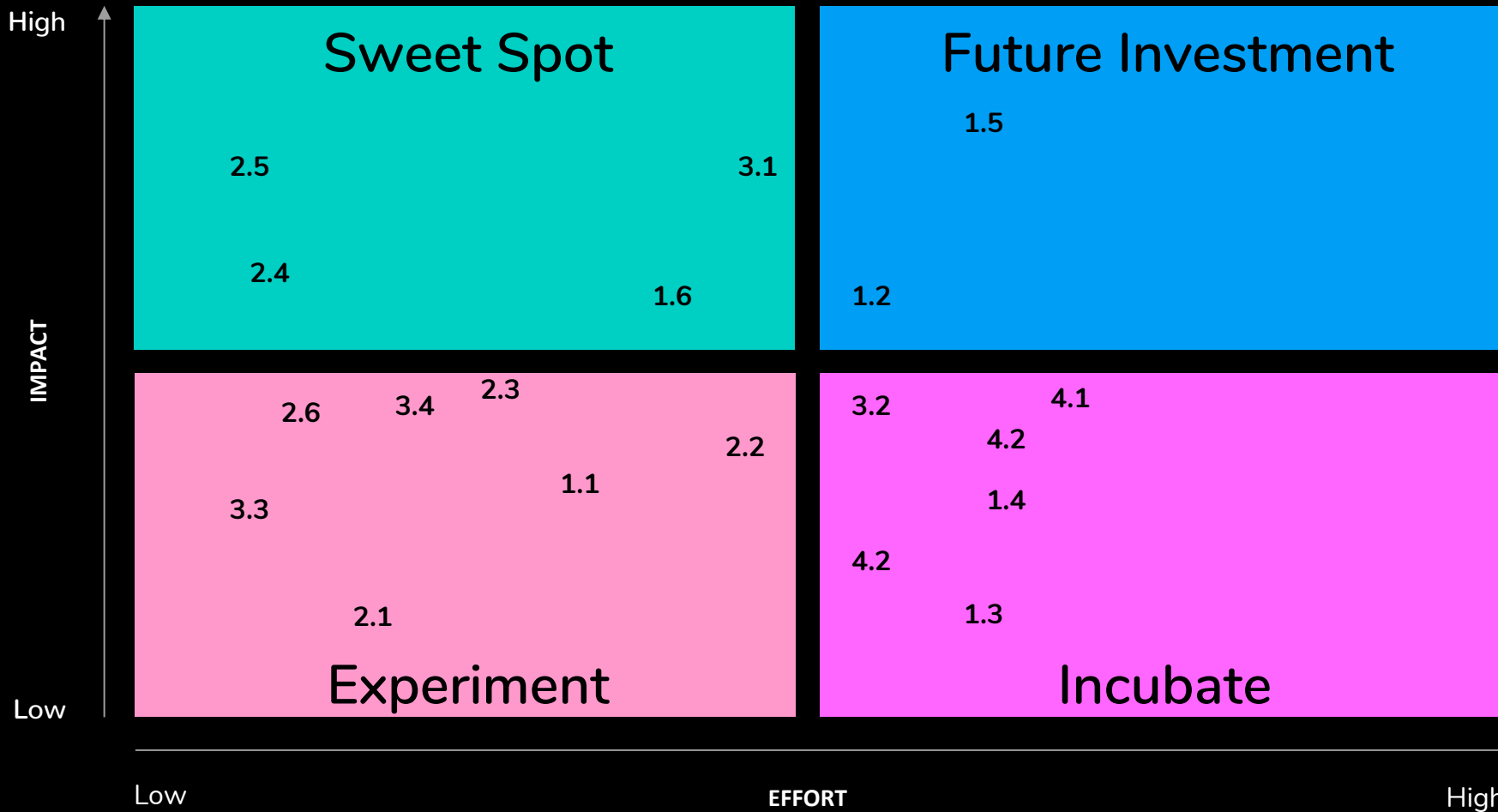


Jodi Blomberg
VP Data Science
Cox Automotive

Diversified Portfolio

- Core Projects ROI Baseline
- Strategic Bets
- Moonshots (4-6 Weeks)

● A Framework



HYPOTHETICAL SCENARIO

AI Use Cases for an Energy company, segmented by target initiative:

1. Productivity

1. Flow Rate Prediction
2. Downhole Pressure Optimization
3. Breccia Image Classification
4. Mean Grain Size Estimation
5. Well Segmentation
6. Demand Visibility Prediction

2. Efficiency

1. Equipment Wear & Tear
2. Predictive Maintenance
3. Optimal Pump Down
4. P&ID Drawing Extraction
5. Instrumentation Issue Detection
6. Material Master Search

3. Safety

1. CV Enabled Asset Management
2. Wellhole Prediction
3. Structure Detection Identification
4. Cable Tension Modeling

4. Environment

1. Smart Power Notification
2. Mine Aquifer Seepage
3. Pipeline Integrity



-
1. Trends & Adoption
 2. Build vs Buy
 3. Open Source
 4. Gen AI Design for Enterprise
 5. Case Study Examples

🔔 2024-09-29: arxiv.org is experiencing DB issues. 🔔



We gratefully acknowledge support from the Simons Foundation, [member institutions](#), and all contributors.

Donate

arXiv > cs > arXiv:1706.03762

Search... All fields Search

Help | Advanced Search

Computer Science > Computation and Language

[Submitted on 12 Jun 2017 (v1), last revised 2 Aug 2023 (this version, v7)]

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Comments: 15 pages, 5 figures
Subjects: **Computation and Language (cs.CL)**; Machine Learning (cs.LG)
Cite as: arXiv:1706.03762 [cs.CL]
(or arXiv:1706.03762v7 [cs.CL] for this version)
<https://doi.org/10.48550/arXiv.1706.03762>

Submission history

- From: Llion Jones [view email]
- [v1] Mon, 12 Jun 2017 17:57:34 UTC (1,102 KB)
 - [v2] Mon, 19 Jun 2017 16:49:45 UTC (1,125 KB)
 - [v3] Tue, 20 Jun 2017 05:20:02 UTC (1,125 KB)
 - [v4] Fri, 30 Jun 2017 17:29:30 UTC (1,124 KB)
 - [v5] Wed, 6 Dec 2017 03:30:32 UTC (1,124 KB)
 - [v6] Mon, 24 Jul 2023 00:48:54 UTC (1,124 KB)
 - [v7] Wed, 2 Aug 2023 00:41:18 UTC (1,124 KB)

Access Paper:

- View PDF
- HTML (experimental)
- TeX Source
- Other Formats

view license

Current browse context:

cs.CL

< prev | next >

new | recent | 2017-06

Change to browse by:

cs
cs.LG

References & Citations

- NASAADS
- Google Scholar
- Semantic Scholar

120 blog links (what is this?)

DBLP - CS Bibliography

listing | bibtex

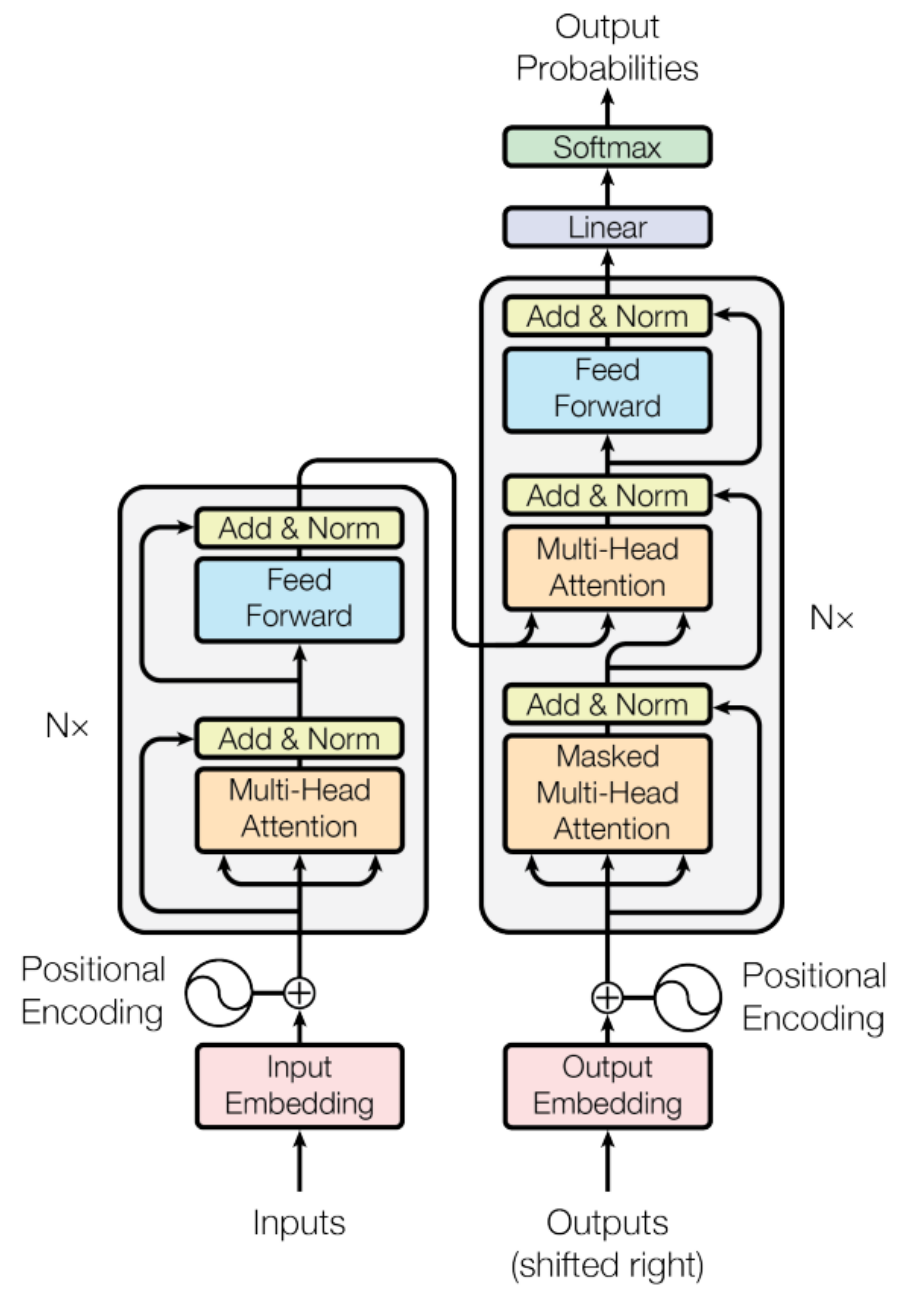
Ashish Vaswani
Noam Shazeer
Niki Parmar
Jakob Uszkoreit
Llion Jones

...

Export BibTeX Citation

Bookmark







15,038

PACKAGE REFERENCE

[Edit on GitHub](#)

Sentence Transformers v3.2 just released, introducing the ONNX and OpenVINO backends for Sentence Transformer models. Read [SentenceTransformer > Usage > Speeding up Inference](#) to learn more about the new backends and what they can mean for your inference speed.

See the [Quickstart](#) for more quick information on how to use Sentence Transformers.

Using Sentence Transformer models is elementary:

```
from sentence_transformers import SentenceTransformer
```

1. Load a pretrained Sentence Transformer model

Installation

SentenceTransformers Document

×

+

←

→

↺

🏠

🔍

sbert.net

☆

👁

N


📁

|

⬇

👤

⋮



SBERT.net

🏠

Sentence Transformers

😊

Star

15,038

Search docs

GETTING STARTED

🏠 » SentenceTransformers Documentation

Edit on GitHub

💡 Tip

Sentence Transformers v3.2 just released, introducing the ONNX and OpenVINO backends for Sentence Transformer models. Read [SentenceTransformer > Usage > Speeding up Inference](#) to learn more about the new backends and what they can mean for your inference speed.

SentenceTransformers Documentation

A wide selection of over 5,000 pre-trained Sentence Transformers models are available for immediate use on 😊 Hugging Face, including many of the state-of-the-art models from the Massive Text Embeddings Benchmark (MTEB) leaderboard. Additionally, it is easy to train or finetune your own models using Sentence Transformers, enabling you to create custom models for your specific use cases.

Models - Hugging Face

huggingface.co/models

Hugging Face

Search models, datasets, users...

Models Datasets Spaces Posts Docs Solutions Pricing Log In Sign Up

Tasks Libraries Datasets Languages Licenses Other

Filter Tasks by name

Multimodal

Image-Text-to-Text Visual Question AnsweringDocument Question Answering Video-Text-to-TextAny-to-Any

Computer Vision

Depth Estimation Image ClassificationObject Detection Image SegmentationText-to-Image Image-to-Text Image-to-ImageImage-to-Video Unconditional Image GenerationVideo Classification Text-to-VideoZero-Shot Image Classification Mask GenerationZero-Shot Object Detection Text-to-3DImage-to-3D Image Feature ExtractionKeypoint Detection

Natural Language Processing

Text Classification Token ClassificationTable Question Answering Question AnsweringZero-Shot Classification Translation

Models 1,050,361 Filter by name

Full-text search Sort: Trending

rain1011/pyramid-flow-sd3

Text-to-Video Updated 3 days ago 470

rhymes-ai/Aria

Image-Text-to-Text Updated 1 day ago 7.56k 332

SWivid/F5-TTS

Text-to-Speech Updated 1 day ago 262

nvidia/NVLM-D-72B

Image-Text-to-Text Updated 5 days ago 23k 620

jxm/cde-small-v1

Feature Extraction Updated 4 days ago 2.82k 219

stepfun-ai/GOT-OCR2_0

Image-Text-to-Text Updated 26 days ago 208k 1.03k

meta-llama/Llama-3.2-11B-Vision-Instruct

Image-Text-to-Text Updated 14 days ago 733k 634

alimama-creative/FLUX.1-Turbo-Alpha

Text-to-Image Updated about 2 hours ago 554 72

meta-llama/Llama-3.2-3B-Instruct

openai/whisper-large-v3-turbo

Automatic Speech Recognition Updated 10 days ago 140k 954

black-forest-labs/FLUX.1-dev

Text-to-Image Updated Aug 16 1.08M 5.44k

ostris/OpenFLUX.1

Text-to-Image Updated 10 days ago 14.6k 483

apple/DepthPro

Depth Estimation Updated 4 days ago 220

meta-llama/Llama-3.2-1B

Text Generation Updated 13 days ago 258k 469

black-forest-labs/FLUX.1-schnell

Text-to-Image Updated Aug 16 1.05M 2.57k

arcee-ai/SuperNova-Medius

Text Generation Updated 3 days ago 945 79

meta-llama/Llama-3.1-8B-Instruct

Text Generation Updated 18 days ago 2.71M 2.77k

allenai/Molmo-7B-D-0924

https://huggingface.co/black-forest-labs/FLUX.1-dev

all-MiniLM-L6-v2

- Trained on over 1 billion samples
- State-of-the-art sentence embeddings (x384)
- Has 6 transformer encoder layers
- A small model
- Mimics performance of large BERT model
- Use-cases: Text classification, sentiment analysis, semantic search


```
# to read and manipulate the data
import pandas as pd
import numpy as np
pd.set_option('max_colwidth', None)    # setting column to the maximum column width as per the data

# to visualise data
import matplotlib.pyplot as plt
import seaborn as sns

# to compute distances
from scipy.spatial.distance import cdist, pdist
from sklearn.metrics import silhouette_score

# importing the PyTorch Deep Learning library
import torch

# to import the model
from sentence_transformers import SentenceTransformer

# to cluster the data
from sklearn.cluster import KMeans

# to compute metrics
from sklearn.metrics import classification_report

# to avoid displaying unnecessary warnings
import warnings
warnings.filterwarnings("ignore")
```

› Loading the dataset

NLP7_CategorizeNewsArticles_

colab.research.google.com/drive/14YLCDmKAzeGSV_TzZCFh2FMcp80dF4hW#scrollTo=yNDAvKn0Klwa

☆

🔗

📄

📥

👤

⋮

CO NLP7_CategorizeNewsArticles_CaseStudy.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at October 12

Comment Share ⚙️ 👤

+ Code + Text

Connect T4 Gemini

⬆️ ⬇️ 🔗 💬 ✎️ 📄 🗑️ ⋮

▶️ #Defining the model

model = SentenceTransformer('sentence-transformers/all-MiniLM-L6-v2')

🔄

.gitattributes: 100% 1.18k/1.18k [00:00<00:00, 83.4kB/s]

1_Pooling/config.json: 100% 190/190 [00:00<00:00, 11.5kB/s]

README.md: 100% 10.6k/10.6k [00:00<00:00, 660kB/s]

config.json: 100% 612/612 [00:00<00:00, 49.0kB/s]

config_sentence_transformers.json: 100% 116/116 [00:00<00:00, 9.65kB/s]

data_config.json: 100% 39.3k/39.3k [00:00<00:00, 2.74MB/s]

pytorch_model.bin: 100% 90.9M/90.9M [00:00<00:00, 253MB/s]

sentence_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 3.09kB/s]

special_tokens_map.json: 100% 112/112 [00:00<00:00, 9.66kB/s]

tokenizer.json: 100% 466k/466k [00:00<00:00, 2.42MB/s]

tokenizer_config.json: 100% 350/350 [00:00<00:00, 27.5kB/s]

train_script.py: 100% 13.2k/13.2k [00:00<00:00, 937kB/s]

vocab.txt: 100% 232k/232k [00:00<00:00, 9.17MB/s]

modules.json: 100% 349/349 [00:00<00:00, 15.4kB/s]

<> > Encoding the dataset

[] 6 cells hidden

> Semantic Search


```
# defining the query text
query_text = "Budget for elections"

# displaying the top 5 similar sentences
top_k_reviews = top_k_similar_sentences(embedding_matrix, query_text, 5)

for i in top_k_reviews:
    print(i, end="\n\n")
```

A cap on donations to political parties should not be introduced yet, the elections watchdog has said.

Fears that big donors can buy political favours have sparked calls for a limit. In a new report, the Electoral Commission says it is worth debating a Â£10,000 cap for the future but

There were almost Â£68m in reported donations to political parties in 2001, 2002 and 2003, with nearly Â£12m of them from individual gifts worth more than Â£1m. The rules have already

Even without changes the commission does urge political parties to seek out more small-scale donations and suggests there should be income tax relief for gifts under £200. It also

The report suggests doubling the money which can be spent by candidates, while cutting national spending limits from £20m to £15m. The commission also says the spending limits for

+ Code + Text

Connect T4 ▼ ✨ Gemini ▲

```
[ ] # defining the query text
query_text = "High imports and exports"

# displaying the top 5 similar sentences
top_k_reviews = top_k_similar_sentences(embedding_matrix, query_text, 5)

for i in top_k_reviews:
    print(i, end="\n\n")
```

➡ US trade gap ballooned in October

Trade gap narrows as exports rise

The country's surplus on trade-in-services remained steady at Â£1.5bn for the fifth month in a row. Paul Dales, UK economist for Capital Economics, said the figures represented an imp

US trade gap hits record in 2004

NLP7_CategorizeNewsArticles_C

colab.research.google.com/drive/14YLCDmKAzeGSV_TzZCFh2FMcp80dF4hW#scrollTo=3hm4-zm4j8vb

☆

N

CO

NLP7_CategorizeNewsArticles_CaseStudy.ipynb

☆

File Edit View Insert Runtime Tools Help Last saved at October 12

Comment

Share

+ Code + Text

Connect T4 Gemini

Comparing with Actual Categories

[] # loading the actual labels
labels = pd.read_csv("/content/drive/news_article_labels.csv")

[] labels.shape

(2127, 1)

[] # checking the unique labels
labels['Label'].unique()

array(['Politics', 'Sports', 'Technology', 'Entertainment', 'Business'],
dtype=object)

[] # adding the actual categories to our dataframe
clustered_data['Actual Category'] = labels['Label'].values

print(classification_report(clustered_data['Actual Category'], clustered_data['Category']))

	precision	recall	f1-score	support
Business	0.96	0.93	0.95	503
Entertainment	0.96	0.95	0.95	369
Politics	0.96	0.95	0.95	403
Sports	0.98	0.99	0.99	505
Technology	0.91	0.97	0.93	347
accuracy			0.96	2127
macro avg	0.95	0.96	0.96	2127
weighted avg	0.96	0.96	0.96	2127



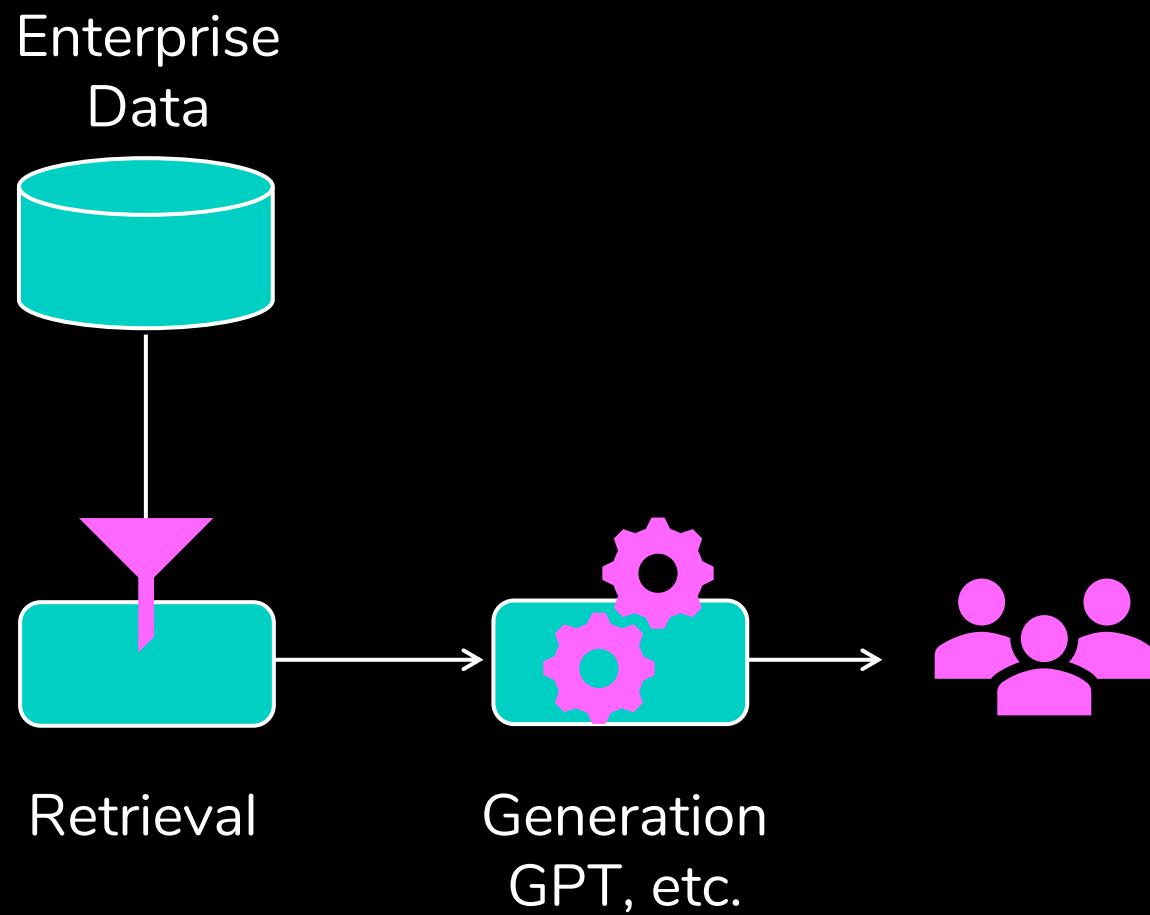
-
1. Trends & Adoption
 2. Build vs Buy
 3. Open Source
 4. Gen AI Design for Enterprise
 5. Case Study Examples

● Enterprise Context

Retrieval Augmented Generation (RAG)

- Combine the strengths of retrieval and generative models.
- More coherent and contextually-relevant responses.

● Enterprise Context



● Knowledge Graph



A way to organize and connect information in a meaningful manner.

- Entities (people, places, events)
- Properties
- Relationships

● Database



Enables similarity-based retrieval.

RAG can leverage this to retrieve similar documents or knowledge from your database, contributing to more informed and contextually relevant generation.

● Database

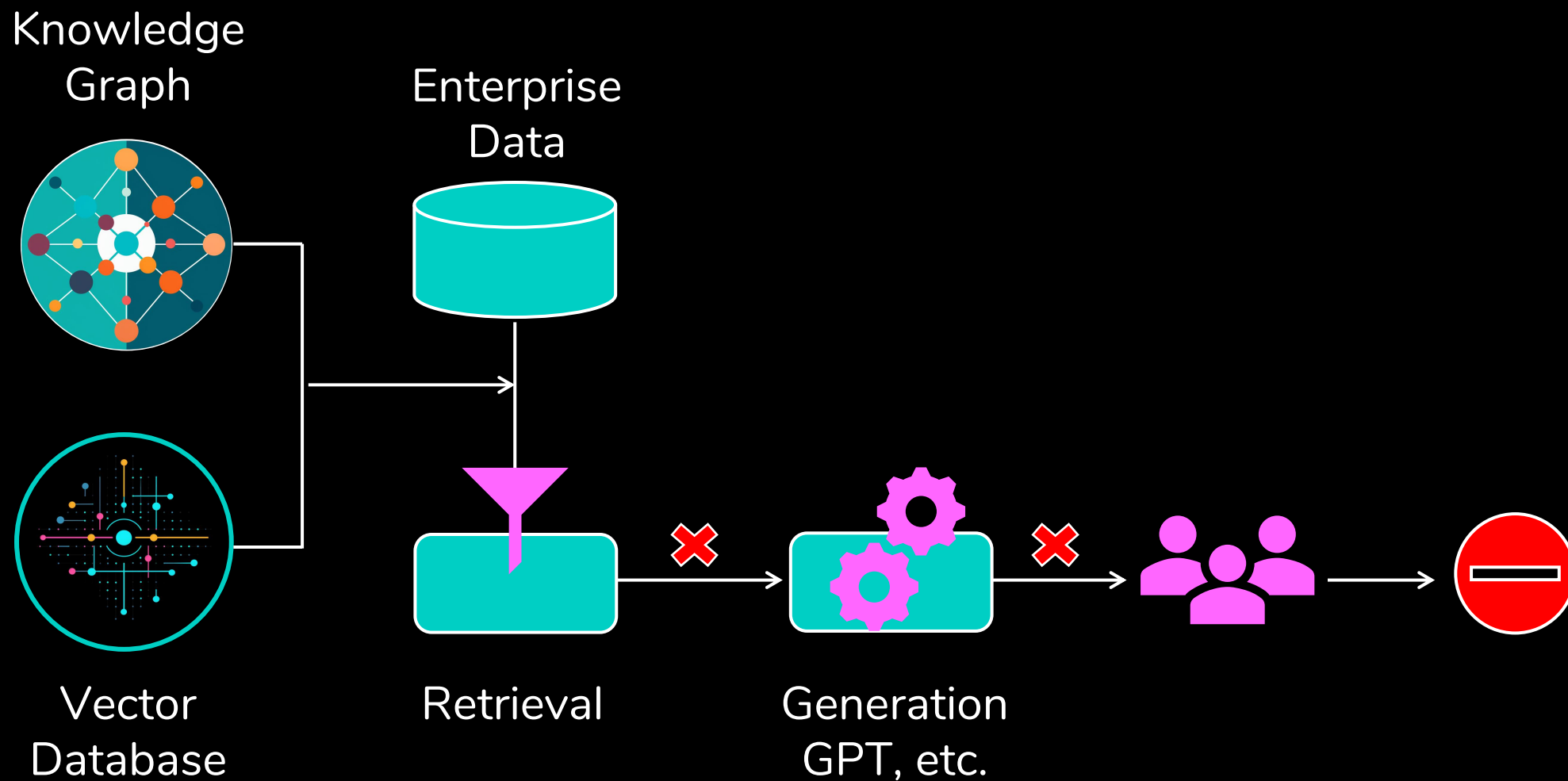


- Apache Airflow
- Apache NiFi
- Talend
- Dataiki DSS
- Data Factory (Azure)
- Glue (aws)

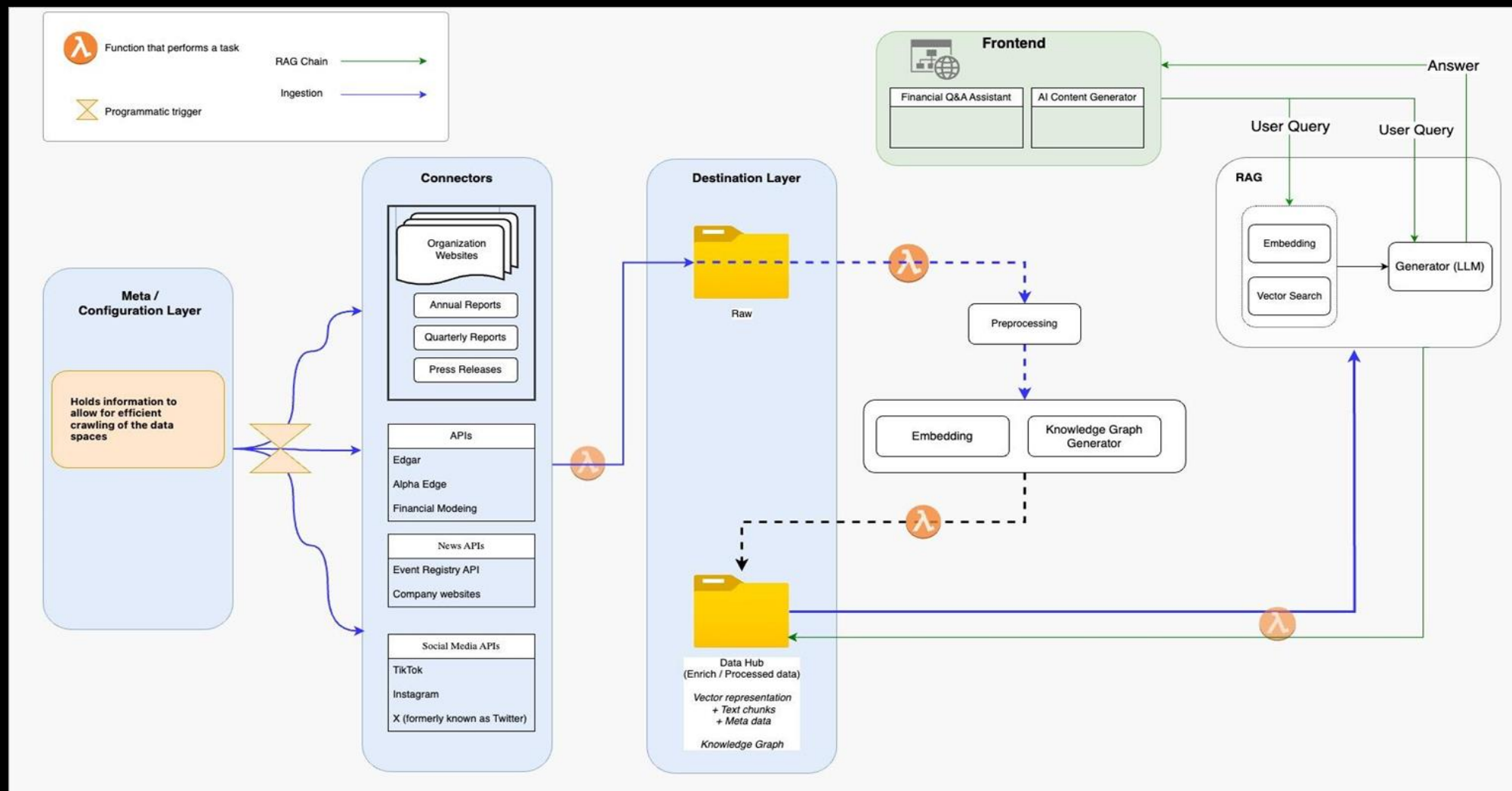
● Data Security

- Private Environment. Customization
- Data Isolation & Encryption
- Access Control. Audit Trails. Monitoring
- Redaction and Anonymization
- Training Data Review

● Full Enterprise Context



● Sample Architecture





-
1. Trends & Adoption
 2. Build vs Buy
 3. Open Source
 4. Gen AI Design for Enterprise
 5. Case Study Examples

AI MASTER

GROUP



Celia Wanderley
Chief Innovation Officer
Bits in Glass



Jim Griffin







Search



Home



My Network



Jobs



Messaging



Notifications



Me

bsi.



Inspiring trust for a more resilient world



Gary Robinson ✓ · 1st

Commercial Solutions Director at BSI Group

Ottawa, Ontario, Canada · [Contact info](#)

500+ connections



Syed Abbas, Srinidhi Rao, and 16 other mutual connections

Message

Save in Sales Navigator

More



BSI



Massachusetts Institute of Technology

AI MASTER

— GROUP —



Ivan Pinto
AVP Delivery
Robosoft Technologies



Jim Griffin





GitHub
Copilot



Code
Whisper

Time Savings

40%

30%

Quality Improvement

10%

20%

AI MASTER

GROUP



Ramsu Sundararajan, PhD
Head of R&D
solus.ai



Jim Griffin





Neil Nathan ✓

Project Scientist at the Benioff Ocean Science Laboratory

Santa Barbara, California, United States · [Contact info](#)

491 connections

[✉ Message](#)

[Save in Sales Navigator](#)

[More](#)



**Benioff Ocean Science
Laboratory**



Stanford University

About

Scientific researcher with a background in biology, marine conservation, and environmental policy. My interests include ecosystem services approaches, global marine policy, conservation project management, and urban sustainability. I have experience working across scales, from local to international, on topics ranging from marine protected areas to natural capital assessments/approaches to the intersections of nature and health. I also have demonstrated technical e ...see more

Google
for
Startups

ultralytics
YOLO Vision

AI MASTER

GROUP



Iqbal Hossain
Director, Research Data Science
University of Arizona



Jim Griffin



Explore research activities

Search for experts, explore research areas, and find collaborations at the University of Arizona using our Knowledge Map (KMAP).

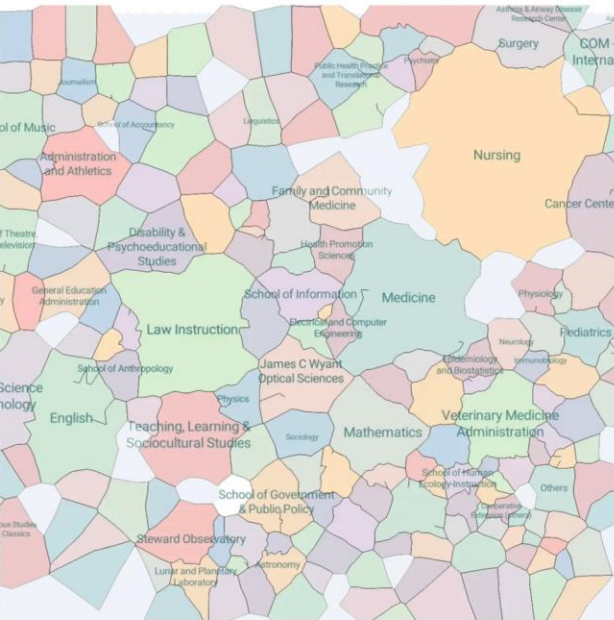
GO TO MAP

See how to explore the university research map

Ask question

ASK KMAP

Get information about the university from our generative AI



Research at a glance

Statistics derived from only the active researchers within the university

RESEARCHERS

6056

Active researchers in KMap

CONNECTIONS

25.6K

Internal connections between researchers

DEPARTMENTS

252

Departments in KMap

GRANTS

23K

Grants for active and completed projects in KMap

CITATIONS

6.5M

Research citations



-
- Voice-activated workflows
 - Semantic search, text classification
 - Code assistant
 - MLOps
 - Computer Vision
 - Knowledge Map



Jim Griffin

Faculty, UT Austin

jim@aimast.org

LinkedIn [jimgriffin4ai](#)

Podcast [AI Master Group](#)