

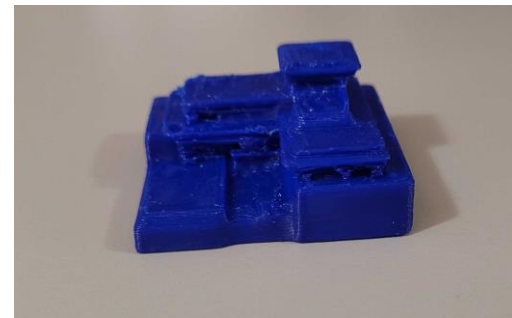


Keboola

# Reflections from the AI Integration Frontline

# About Me

- Research Lead; AI Applications
- Exploring and implementing AI technologies in data pipelines
- Addressing GenAI disruption in data and analytics

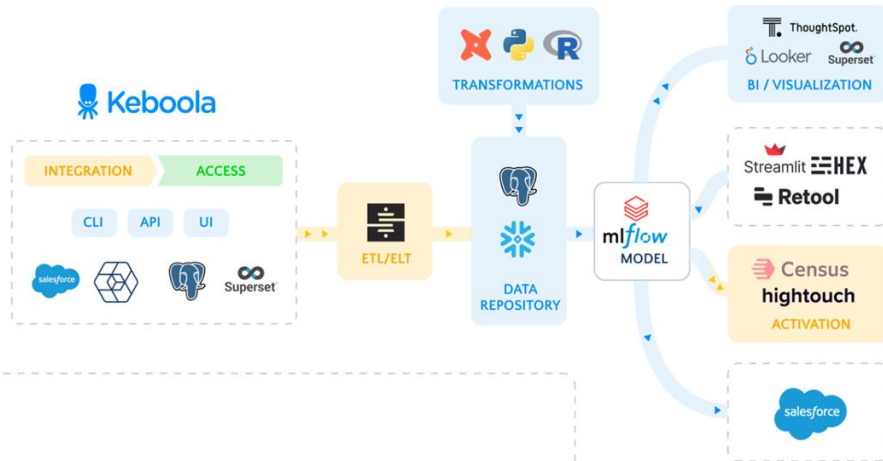


'A house in the style of Frank Lloyd Wright's Fallingwater'

Generated with Point-E from OpenAI



# What is Keboola?



## Orchestration

EXTRACT

LOAD

TRANSFORM

ACTIVATE

### Workflow

- ✓ Execution
- ✓ Scheduling
- ✓ Failover
- ✓ Distribution
- ✓ Verification

### Monitoring

- ✓ Performance
- ✓ Logging
- ✓ Debugging
- ✓ Health Check
- ✓ Scheduling

### Governance & Security

- ✓ User roles access & rights
- ✓ Metadata for security
- ✓ FinOps

# A View of the AI Hype Cycle from the Factory Floor:



**The peak of inflated  
expectations**

Generative AI



**The Trough of  
Disillusionment**

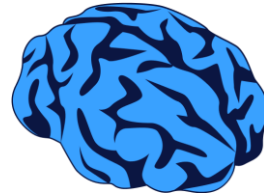
# Two Contrasting Terms Collide:

A decorative graphic on the left side of the slide, consisting of a black wavy line with a series of white circles of varying sizes along its length, resembling an octopus tentacle.

**AI**

**Artificial  
Intelligence**

+



**Natural  
Stupidity**

A decorative graphic on the right side of the slide, consisting of a black wavy line with a series of white circles of varying sizes along its length, resembling an octopus tentacle.

# Struggling with Outliers;

## Unable to handle genuine novelty



How many times does the letter "i" appear in "artificial"?



The letter "i" appears twice in the word "artificial."



## Struggling with Outliers;

Unable to handle  
genuine novelty



You

A man and his mother are in a car accident. The mother sadly dies. The man is rushed to the ER. When the doctor sees him, he says, "I can't operate on this man. He's my son!"

How is this possible?



ChatGPT

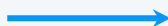
The doctor is the man's other parent—his mother, indicating that the doctor is a woman. This riddle plays on common assumptions about professions and gender roles.

Credit: @colin\_fraser

# Struggling with Outliers;

## Unable to handle genuine novelty

Employee Salaries	
Employee ID	Salary
12	500 CZK/Month
116	1.000.000\$/year
1243	\$75,000/year
242344	3.000/month
555	250.00 Kc/Year
63465	100.000 USD/month



Employee Salaries			
Employee ID	Currency	Salary	Pay Frequency
12	USD	\$5,000	Annual
116	USD	\$1,000,000	Annual
1243	USD	\$75,000	Annual
242344	USD	\$360,000	Annual
555	USD	\$32,000	Annual
63465	USD	1,200,00	Annual



# But where's the value?



**The technology isn't designed to solve the complex problems that would justify the costs**

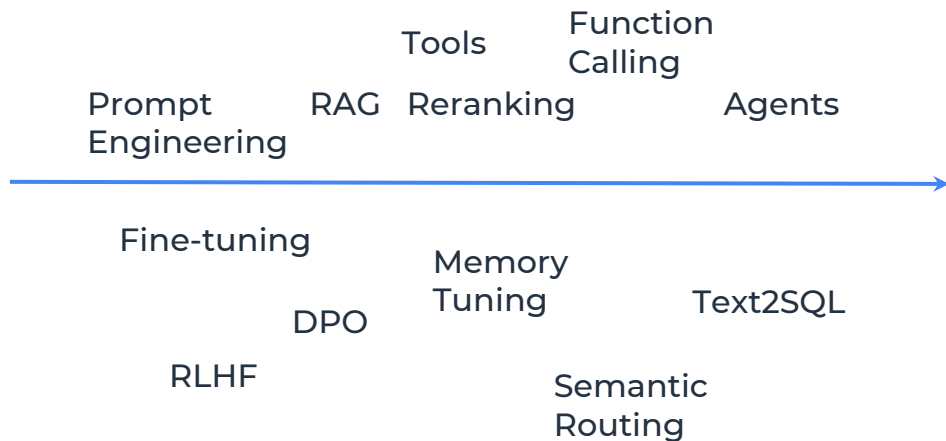
**Goldman Analyst Jim Covello**

# But where's the value?



Investment in AI has reached a new high with a focus on generative AI, which, in most cases, has yet to deliver its anticipated business value.

Gartner - Hype Cycle for Artificial Intelligence, 2024



# But where's the value?

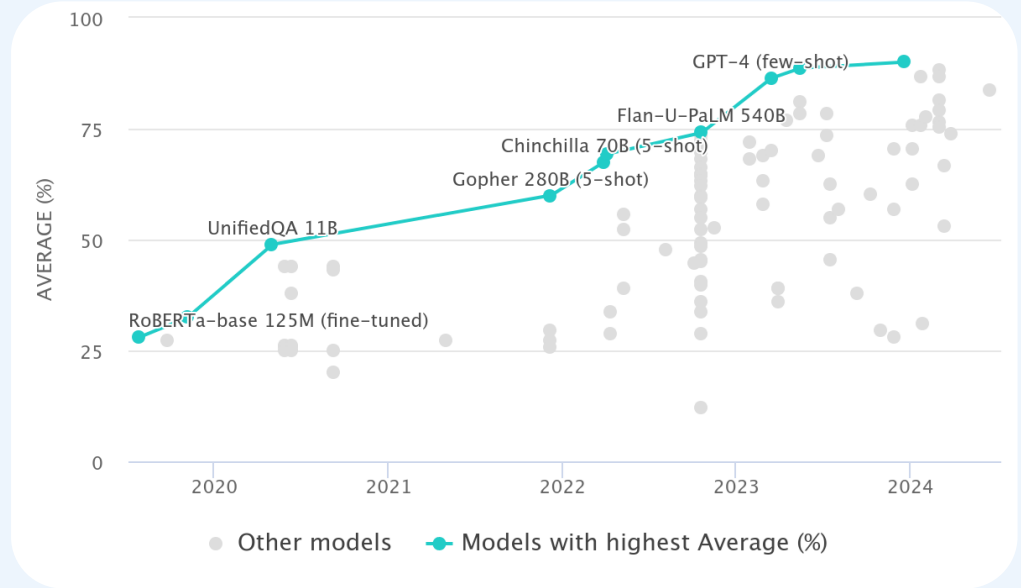
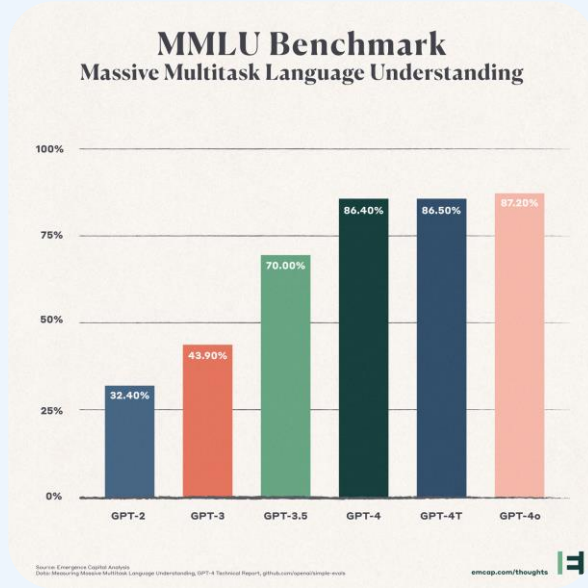


The revenue isn't there yet, and might never come.

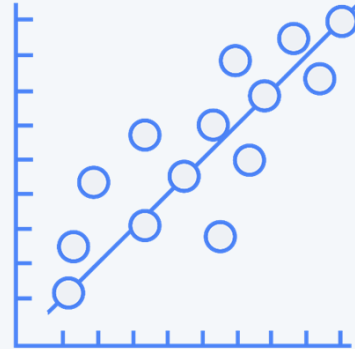
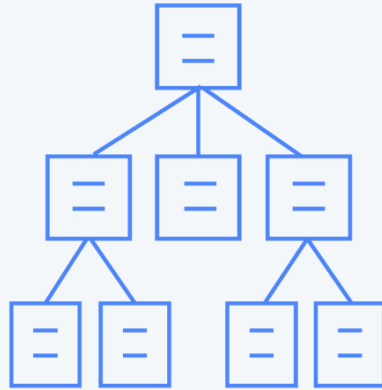
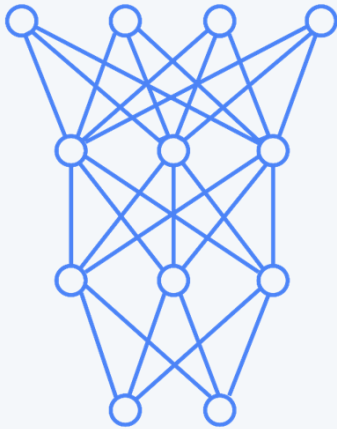
Gary Marcus

All of the applied research going on can be boiled down to one thing: “These models are cool; now how can I get them to do what I actually need?”

# The Plateau is here



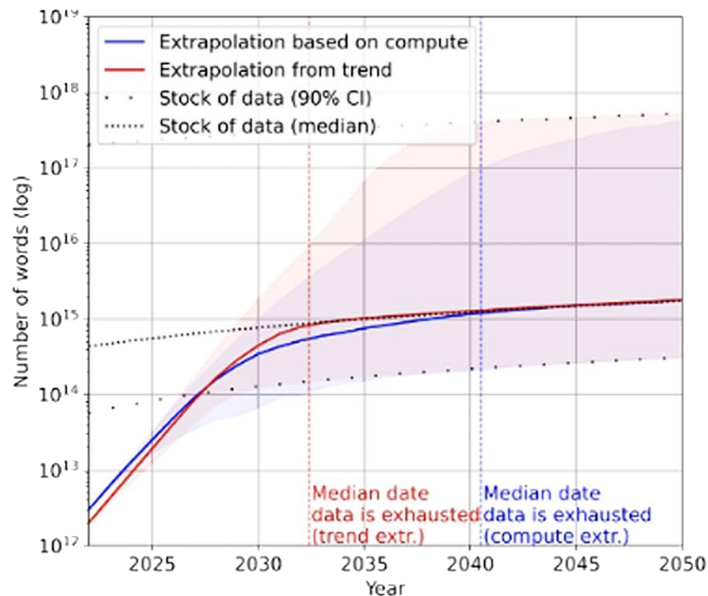
# Data is Central to AI Directly in Machine Learning & Deep learning





# Data is Central to AI Training Foundation Models

## We will run out of data? An analysis of the limits of scaling datasets in Machine Learning

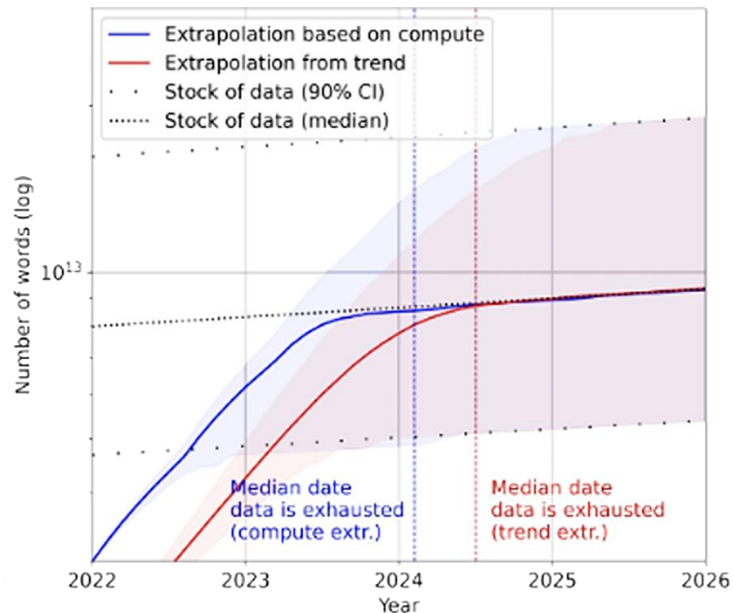


Projections of Low-Quality Data



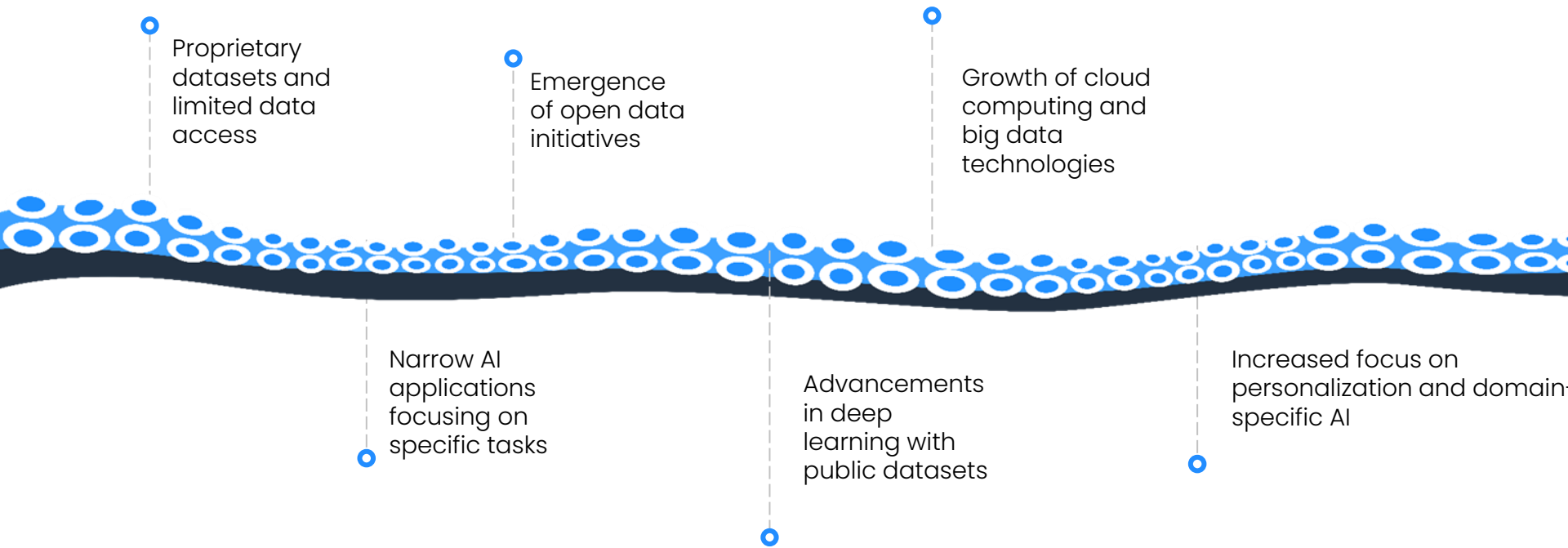
# Data is Central to AI Training Foundation Models

## We will run out of data? An analysis of the limits of scaling datasets in Machine Learning



Projections of High-Quality Data

# From Private Data to Public Data back to Private Data





# Back to “Broccoli AI”

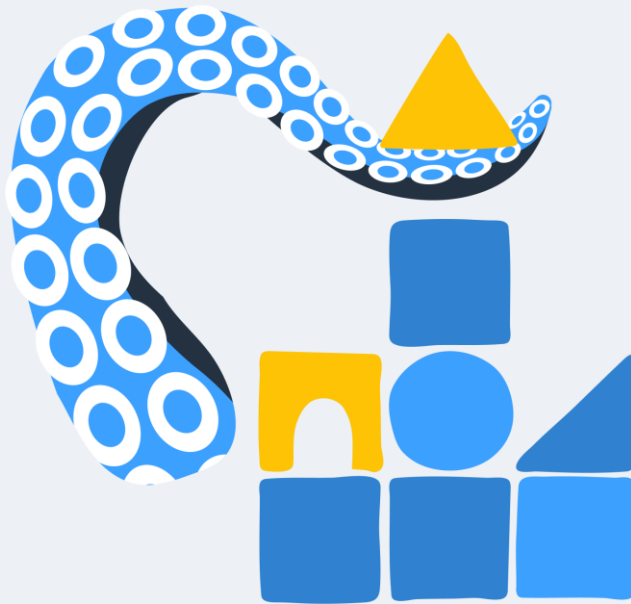
- Demetrios Brinkmann

Aka  
‘Boring AI’



Once we admit that even the current version of frontier models are **not entirely capable of everything** we had hoped for, we are left with two options:

- Do nothing as we wait for the next generation of models
- Sit with these acknowledgements and use them to our advantage to be the “cooler” heads that prevail





## Job Error



SQL Error: Type mismatch: Cannot concatenate a string with an array

### AI Explanation

This error is caused by trying to concatenate a string with an array. **You should first convert the array to a string and try again.** You can use the ARRAY\_TO\_STRING function to resolve this issue.

Is this AI-generated explanation useful?



COPY DETAILS TO CLIPBOARD



CONTACT SUPPORT



ate a string with an array

This error occurs because the SQL query attempts to concatenate a string with an array, which is not supported. **The root cause is a deliberate change made recently in an upstream dependency, which altered the data type of the column being concatenated.**

### Root Cause Analysis:

- **Lineage Tracking:** This error traces back to a recent update in the `user_data_transformation` step, where the `user_interests` column was changed from a string to an array. This change affects the current query in the `report_generation_transformation` step, where the `user_interests` column is concatenated with other strings.

Is this AI-generated explanation useful?



COPY DETAILS TO CLIPBOARD



CONTACT SUPPORT

Shown levels:



Direction:



price-change-history

6 columns

Staff Mailing List

Transformation - Snowflake SQL

Geocoding Location Transformation

Transformation - Python

Customers DC Cleanup

Transformation - Python

EXTERNAL

18 columns

gh-customers

EXTERNAL

95 columns

gh-products

EXTERNAL

17 columns

gh-products

EXTERNAL

38 columns

invoices-demo-products-SS

18 columns

staff-newsletter

7 columns

staff-to-remove

2 columns

geocoded\_location

6 columns

most-recent-staff-changes

3 columns

customers\_m\_cleaned

84 columns

Anomalous Closedout Adjustments

Transformation - Snowflake SQL

Staff Mailing List (Custom Fields)

Transformation - Python

Online Gift Card Sales

Transformation - Snowflake SQL

SS-closedout-adjustments

4 columns

staff-newsletter-final

6 columns

online-gift-card-sales

5 columns

Anomalous Closedout Adjustments Email

Transformation - Python



# Building foundations for AI is a delicate balance



# We need to demand **more** from our tools and vendors

- What is the data for? Does it represent what we want?



# We need to demand **more** from our tools and vendors

- What is the data for? Does it represent what we want?
- What data is actually being used? How is it being processed?





# We need to demand **more** from our tools and vendors

- What is the data for? Does it represent what we want?
- What data is actually being used? How is it being processed?
- How do we validate that the data and models do the things that we want it to do?





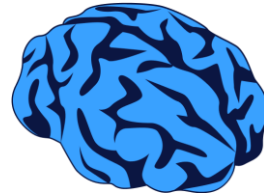
# Two Terms Collide:

A decorative graphic on the left side of the slide, consisting of a black wavy line with a series of white circles of varying sizes along its top edge, resembling a tentacle or a data stream.

**AI**

**Artificial  
Intelligence**

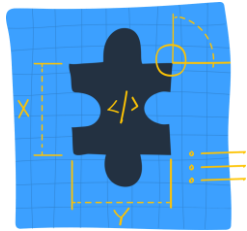
+



**Natural  
Intelligence**

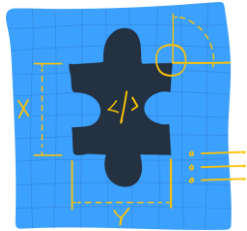
A decorative graphic on the right side of the slide, identical to the one on the left, consisting of a black wavy line with a series of white circles of varying sizes along its top edge, resembling a tentacle or a data stream.

# Lessons from the Trenches



**Be Pragmatic**

# Lessons from the Trenches



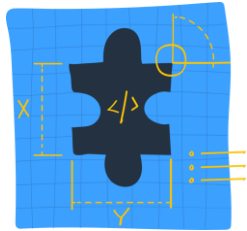
**Be Pragmatic**



**Forge  
Partnerships**



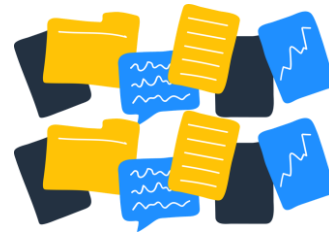
# Lessons from the Trenches



**Be Pragmatic**



**Forge  
Partnerships**



**Curate your  
Knowledge**