

# Generative AI for Government

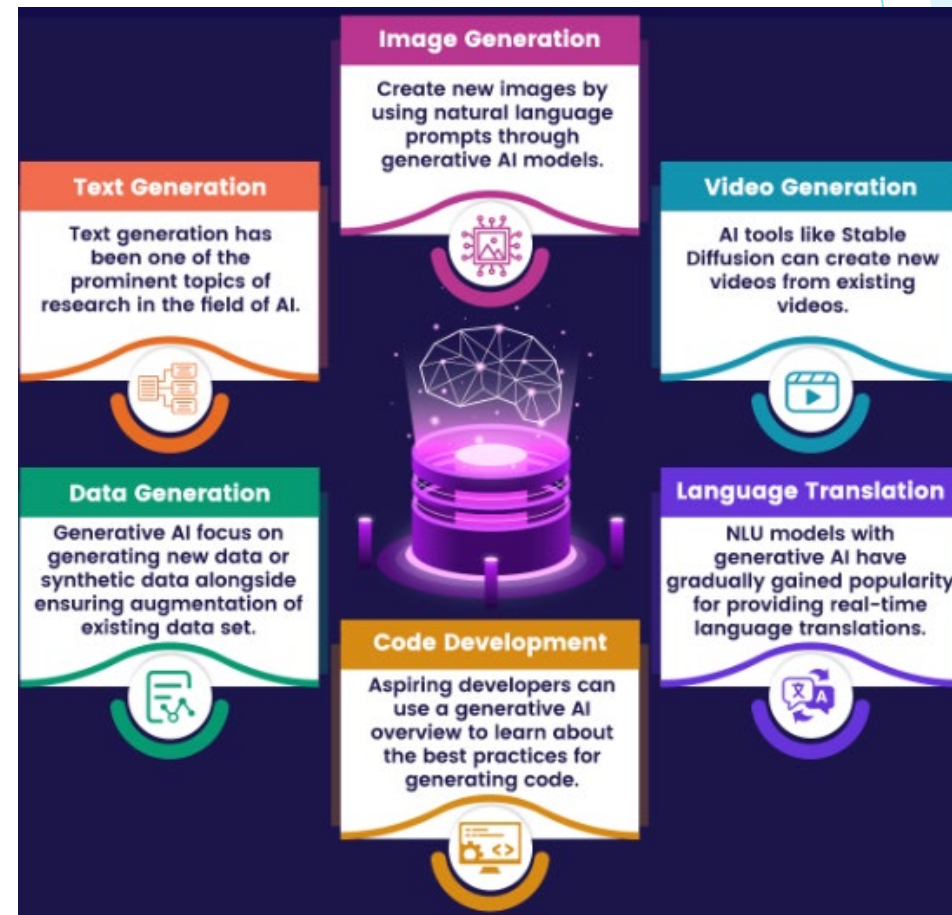
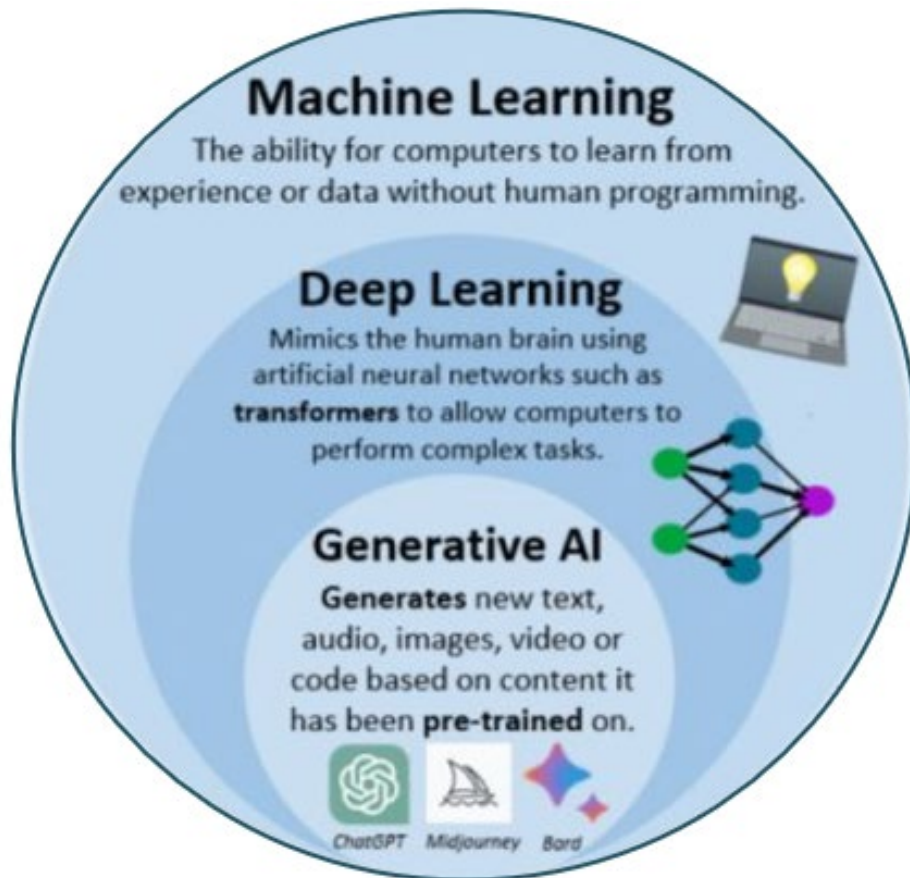
Conner Lawston

Summit2Sea, a cBEYONData Company



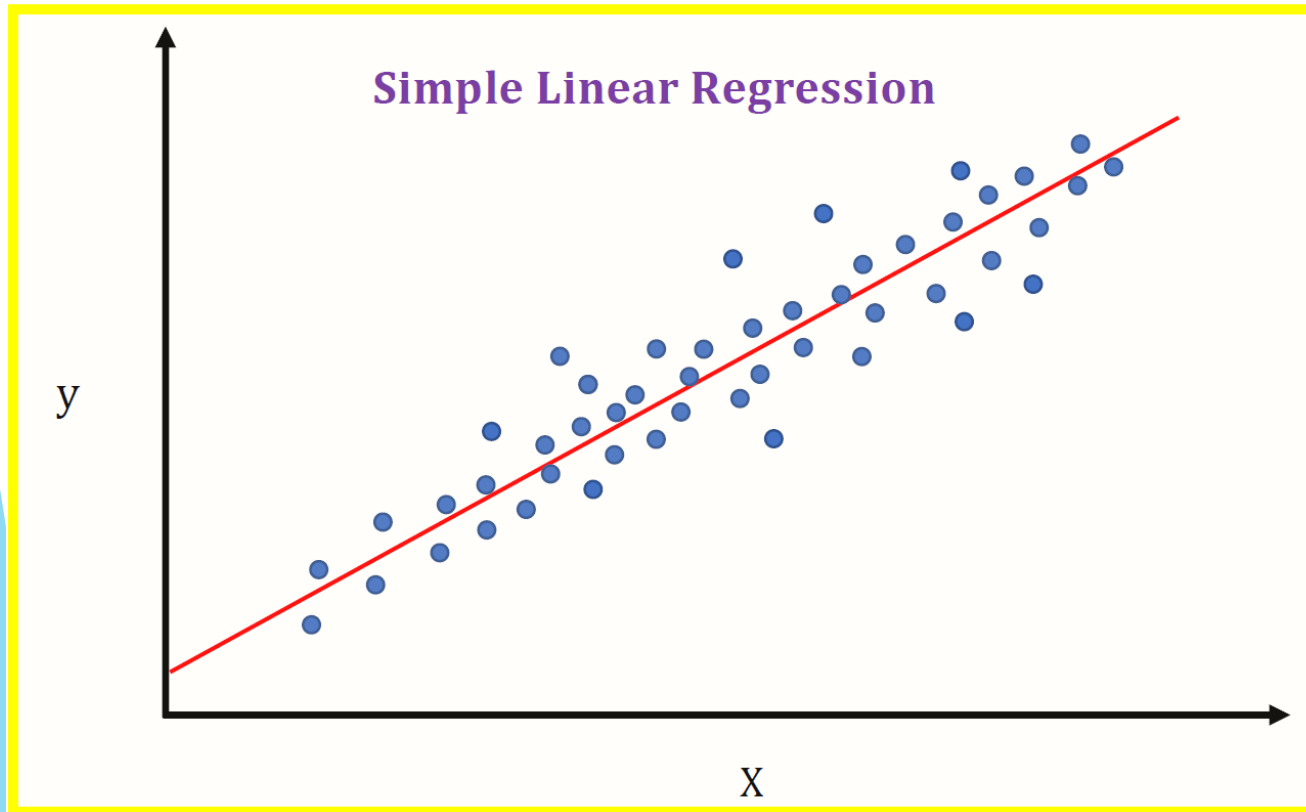
# What is Generative AI?

- ▶ Subset of AI/ML focused on generating new text, images, data, and more
- ▶ It generally uses vast amounts of data (think in terms of billions and trillions of examples)
- ▶ Linear Regression, MUPE, Decision Trees, etc, are not Generative Models- they are used to describe historical patterns



# Comparison- Building Traditional Statistical Models

- ▶ Traditional Statistical modeling involves fitting a model to historical data
- ▶ By doing this we optimize the model Coefficients, or 'Parameters'
- ▶ In ICEAA applications, this parameter optimization is usually done in milliseconds



## CER 1: Statistics

SCOPE			
Primary Methodology	T1, \$K	Bus	Released: November 2020
This CER estimates the T1 cost of a Spacecraft Bus. It does not address the costs for Bus level SEITPM. This CER differs from the other Bus T1 CERs by including as many data points across different types of systems.			
CER			UNITS
$y = 145.63x_1^{0.77} \cdot 0.7^{x_2}$			\$K, FY16
VARIABLES RANGES			
	MIN	MAX	STD DEV
x1: Weight (lbs)	260.00	4,900.00	1,397.94
x2: Agency (1 = Commercial, 0 = Military)	0.00	1.00	0.50
STATS			
PEARSONS R^2			^2 MUPE
0.6892			.7438

**USCM Model:**  
 2 Parameters (Weight, Agency)  
 ~100 data examples (Satellites)

# Comparison- Building Generative AI Models

- ▶ Generative AI models can have millions or billions of Parameters to optimize, requiring trillions of input data points, taking months and millions of dollars to compute!

Model	Billions of Tokens (Compute-optimal)	Days to Train on MosaicML Cloud	Approx. Cost on MosaicML Cloud
GPT-1.3B	26B	0.14	\$2,000
GPT-2.7B	54B	0.48	\$6,000
GPT-6.7B	134B	2.32	\$30,000
GPT-13B	260B	7.43	\$100,000
<b>GPT-30B *</b>	<b>610B</b>	<b>35.98</b>	<b>\$450,000</b>
GPT-70B **	1400B	176.55	\$2,500,000

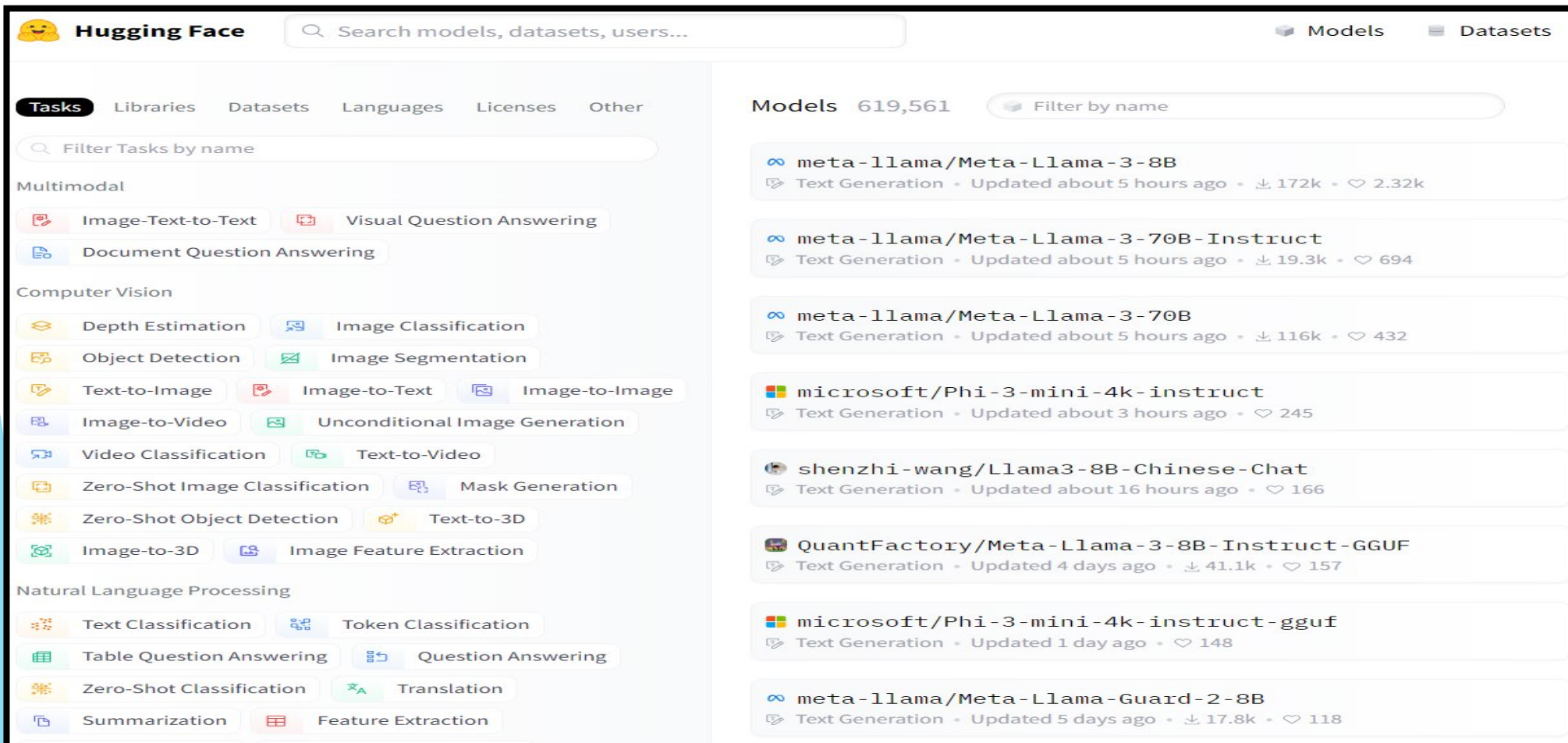
# Building GEN AI Models

- ▶ For example, the company ‘Databricks’ built their Gen AI Text Model, ‘Dolly’ by feeding it 15,000 questions and answers
- ▶ Each one provides ‘context’, a sentence explaining the background/broad subject
- ▶ They are divided into several categories, like summarization or question answering

instruction string · lengths	context string · lengths	response string · lengths	category string · classes
4 11.7k	7.05k=9.4k 0.2%	1 26k	8 values
Given these paragraphs about Large language models, when was the first model in OpenAI's GPT series trained?	A large language model (LLM) is a language model consisting of a neural network with many parameters...	2018.	closed_qa
From the text given, extract the iPhone Models and the announcement dates and list them using the bullets in the...	Jobs unveiled the first-generation iPhone to the public on January 9, 2007, at the Macworld 2007 convention at the...	Here is a bullet list of iPhone Models and the corresponding announcement dates in the format {iPhone...	information_extraction
Using this text, please give me a short bulleted list of famous figures buried at the Sleepy Hollow Cemetery in...	Notable burials Numerous notable people are interred at Sleepy Hollow Cemetery, including: Viola Allen (1867-...	Sleepy Hollow Cemetery in Sleepy Hollow, NY is the final resting place of numerous notable figures, including: -...	summarization
Give me a bulleted list of 3 informal fallacies and the meaning each	Informal fallacies – arguments that are logically unsound for lack of well-grounded premises. Argument to moderatio...	Three informal fallacies are: - False Authority: Using an expert with dubious credentials to promote an argument -...	summarization
Given these paragraphs about Large language models, how many words did the BookCorpus have when it was used to...	A large language model (LLM) is a language model consisting of a neural network with many parameters...	985 million words.	closed_qa
Given these paragraphs about Large language models, what is the prompting paradigm?	A large language model (LLM) is a language model consisting of a neural network with many parameters...	In the prompting paradigm, popularized by GPT-3, the problem to be solved is formulated via a text prompt,...	closed_qa
Given these paragraphs about Large language models, what is an LLM?	A large language model (LLM) is a language model consisting of a neural network with many parameters...	A large language model (LLM) is a language model consisting of a neural network with many parameters...	closed_qa
Provide a summary of Oliver Cromwell Applegate's career and later years.	Captain Oliver Cromwell Applegate (June 11, 1845 – October 11, 1938) was an American politician, newspaper editor,...	Oliver spent his early career as a teacher at the Ashland school as well as a member of the Mountain Rangers. He...	summarization
Given these paragraphs about Large language models, what are some examples of emergent abilities?	A large language model (LLM) is a language model consisting of a neural network with many parameters...	Examples include multi-step arithmetic, taking college-level exams, identifying the intended meaning of a word,...	closed_qa
Given these paragraphs about Large language models, what is hallucination?	A large language model (LLM) is a language model consisting of a neural network with many parameters...	In artificial intelligence in general, and in large language models in particular, a "hallucination" is a...	closed_qa

# Many Competing Models

- ▶ Thousands of different Gen AI models exist, built by hundreds of companies
- ▶ Each use different approaches, datasets, and algorithm architecture
- ▶ Many different tasks, such as Question Answering, Text-to-Video Creation, etc.



The screenshot displays the Hugging Face interface. At the top, there is a search bar for models, datasets, and users. Below the search bar, there are navigation tabs for 'Tasks', 'Libraries', 'Datasets', 'Languages', 'Licenses', and 'Other'. The 'Tasks' tab is selected, and a search bar for filtering tasks by name is visible. The main content area is divided into two columns. The left column lists various tasks categorized into Multimodal, Computer Vision, and Natural Language Processing. The right column shows a list of models with their names, descriptions, and statistics. The models listed include meta-llama/Meta-Llama-3-8B, meta-llama/Meta-Llama-3-70B-Instruct, meta-llama/Meta-Llama-3-70B, microsoft/Phi-3-mini-4k-instruct, shenzhi-wang/Llama3-8B-Chinese-Chat, QuantFactory/Meta-Llama-3-8B-Instruct-GGUF, microsoft/Phi-3-mini-4k-instruct-gguf, and meta-llama/Meta-Llama-Guard-2-8B.

**Hugging Face** Search models, datasets, users... Models Datasets

Tasks Libraries Datasets Languages Licenses Other

Filter Tasks by name

**Multimodal**

- Image-Text-to-Text
- Visual Question Answering
- Document Question Answering

**Computer Vision**

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Text-to-Image
- Image-to-Text
- Image-to-Image
- Image-to-Video
- Unconditional Image Generation
- Video Classification
- Text-to-Video
- Zero-Shot Image Classification
- Mask Generation
- Zero-Shot Object Detection
- Text-to-3D
- Image-to-3D
- Image Feature Extraction

**Natural Language Processing**

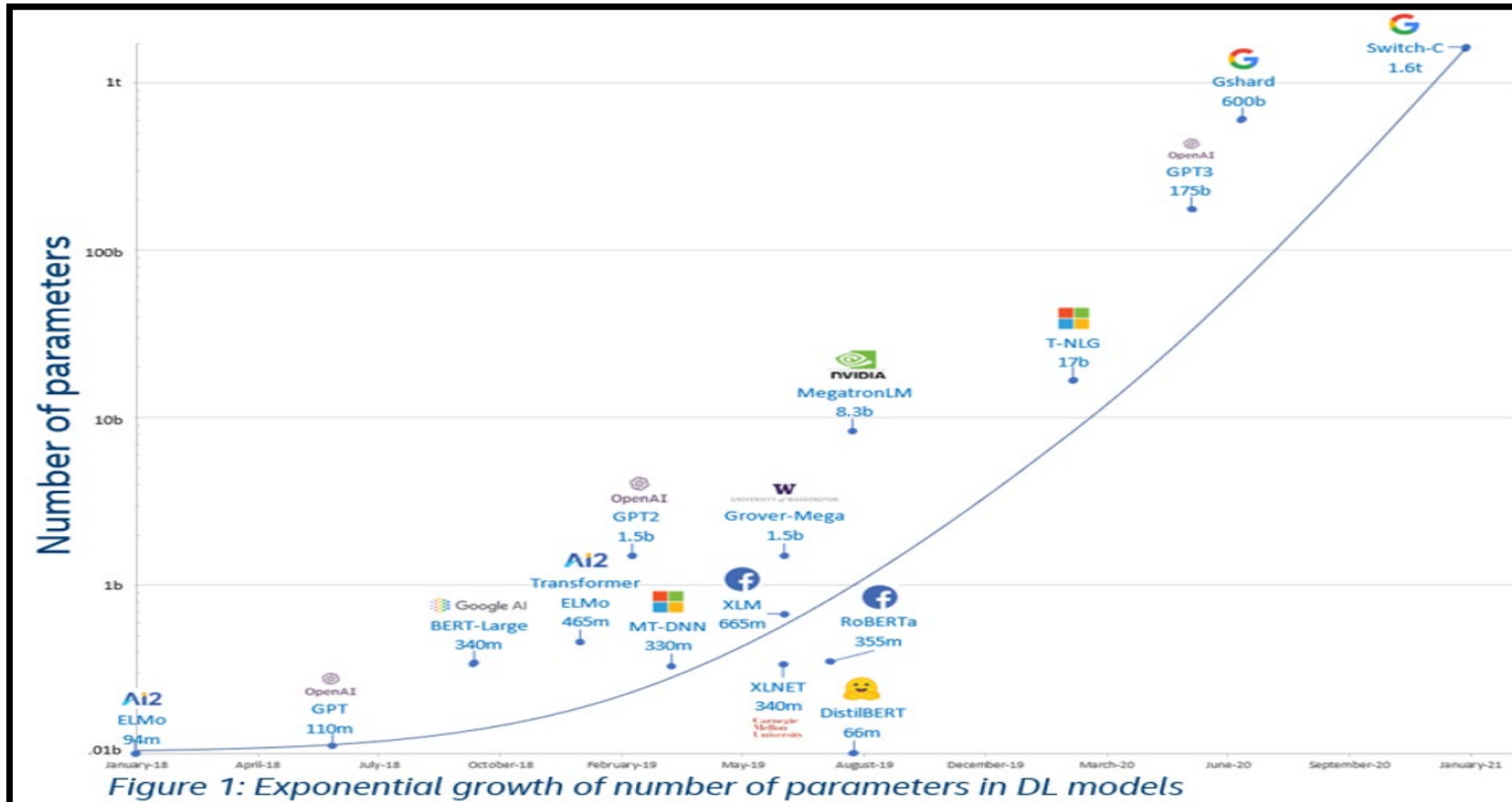
- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Feature Extraction

**Models** 619,561 Filter by name

- meta-llama/Meta-Llama-3-8B  
Text Generation • Updated about 5 hours ago • ↓ 172k • ♥ 2.32k
- meta-llama/Meta-Llama-3-70B-Instruct  
Text Generation • Updated about 5 hours ago • ↓ 19.3k • ♥ 694
- meta-llama/Meta-Llama-3-70B  
Text Generation • Updated about 5 hours ago • ↓ 116k • ♥ 432
- microsoft/Phi-3-mini-4k-instruct  
Text Generation • Updated about 3 hours ago • ♥ 245
- shenzhi-wang/Llama3-8B-Chinese-Chat  
Text Generation • Updated about 16 hours ago • ♥ 166
- QuantFactory/Meta-Llama-3-8B-Instruct-GGUF  
Text Generation • Updated 4 days ago • ↓ 41.1k • ♥ 157
- microsoft/Phi-3-mini-4k-instruct-gguf  
Text Generation • Updated 1 day ago • ♥ 148
- meta-llama/Meta-Llama-Guard-2-8B  
Text Generation • Updated 5 days ago • ↓ 17.8k • ♥ 118

# Sizing Trends of GEN AI Models

- ▶ In general, the models are getting bigger and bigger, and being trained on larger datasets
- ▶ The Original GPT1 in 2018 had 110 million parameters, now some have over a trillion!



# Making Text Understandable with Natural Language Processing (NLP) Techniques

- ▶ Text must be able to be understood by the AI, both when prompting it with instructions to carry out, and in the initial creation/training of the model
- ▶ Stemming- converting words to their root
  - ▶ (“running” -> “run”)
- ▶ Removing “Stop Words” that are meaningless
  - ▶ (“the”, “and”, “a”, etc.)
- ▶ Part of Speech Tagging helps AI understand structure
  - ▶ (“dog”-> noun, “ran” -> verb)
- ▶ “Sentiment Analysis” determines positive or negative connotation
  - ▶ (“delicious food” -> +0.95, “slow service” -> -0.80)

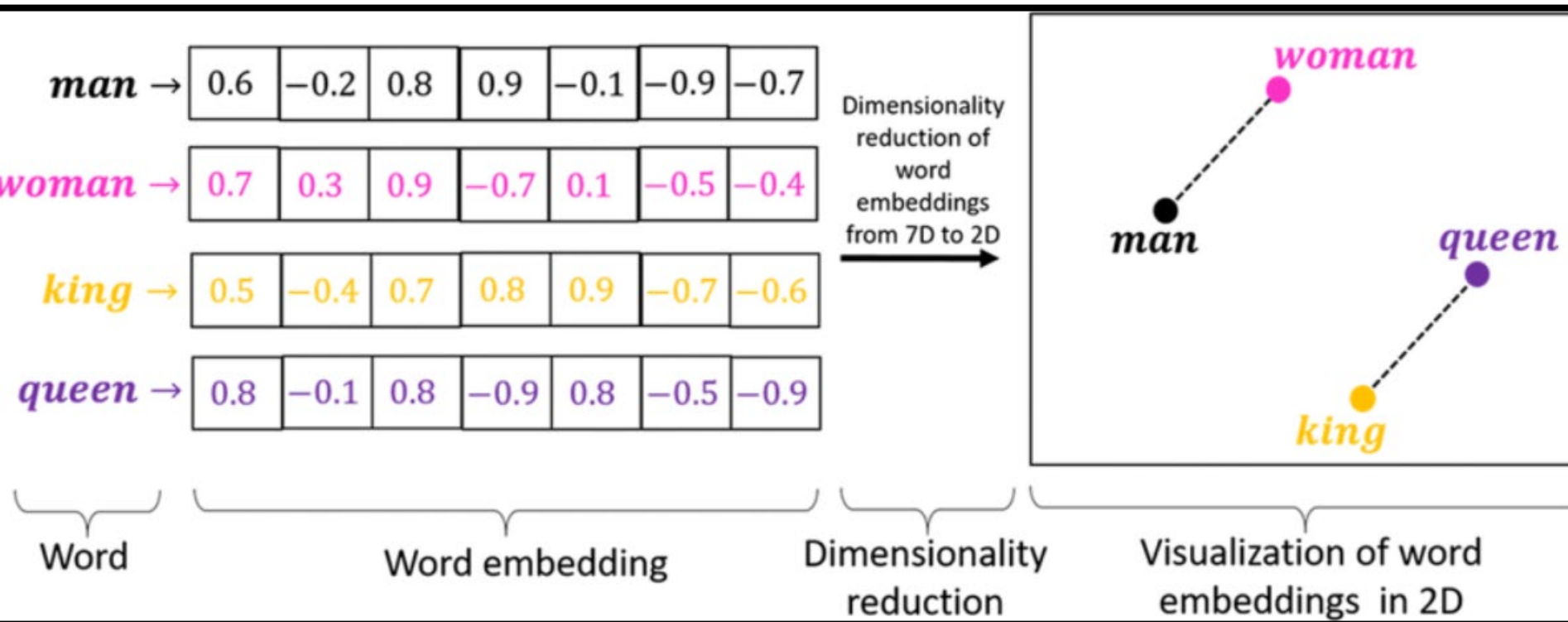




# How GEN AI Models See Data

Presented at the ICEAA 2024 Professional Development & Training Workshop - [www.iceaaonline.com/min2024](http://www.iceaaonline.com/min2024)

- ▶ The textual datasets for GEN AI models are not simply stored as word docs or pdfs, which would be massive and incredibly inefficient (i.e. entire book volumes, all of Wikipedia, etc.)
- ▶ Instead, textual data is converted into a series of numbers, in a process called 'Embedding'
- ▶ The set of numbers, or 'Vector', can now be used for mathematical operations (+, -, x, /, etc.)
- ▶ In this example, 'man' - 'woman' would yield a similar mathematical result to 'king' - 'queen'
- ▶ This process allows the GEN AI models to understand sentence structure, connotation, and more



# Generative AI Applications: Text Generation

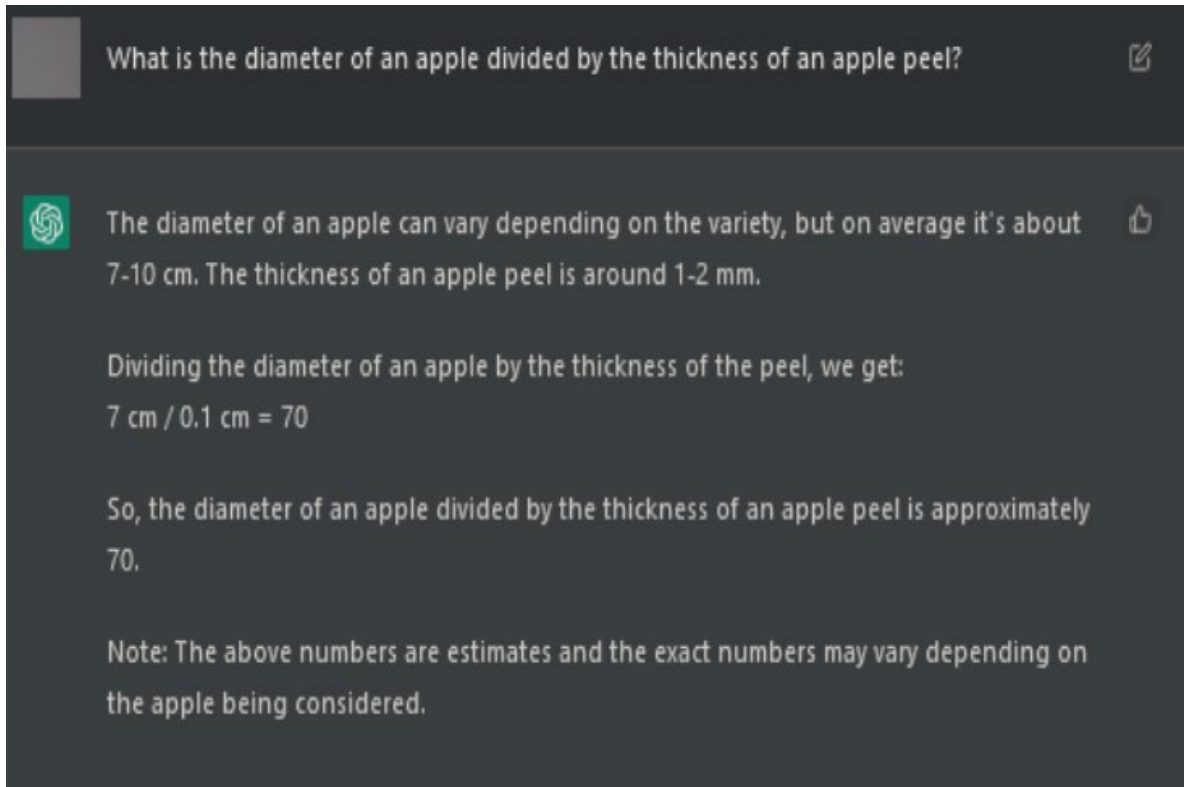
- ▶ Probably the most commonly known GEN AI Text Model (LLM) is OpenAI’s ‘GPT’ Series
- ▶ ‘Generative Pre-Trained Transformer’ (GPT) had its first version released in 2018
- ▶ It gained worldwide attention in 2023 with the release of ‘ChatGPT’ API, based on model 3.5

Model	Architecture	Parameter count	Training data	Release date
GPT-1	12-level, 12-headed Transformer decoder (no encoder), followed by linear-softmax.	117 million	<a href="#">BookCorpus</a> : <sup>[34]</sup> 4.5 GB of text, from 7000 unpublished books of various genres.	June 11, 2018 <sup>[9]</sup>
GPT-2	GPT-1, but with modified normalization	1.5 billion	WebText: 40 GB of text, 8 million documents, from 45 million webpages upvoted on <a href="#">Reddit</a> .	February 14, 2019 (initial/limited version) and November 5, 2019 (full version) <sup>[35]</sup>
GPT-3	GPT-2, but with modification to allow larger scaling	175 billion <sup>[38]</sup>	499 billion tokens consisting of <a href="#">CommonCrawl</a> (570 GB), WebText, English Wikipedia, and two books corpora (Books1 and Books2).	May 28, 2020 <sup>[36]</sup>
GPT-3.5	Undisclosed	175 billion <sup>[38]</sup>	Undisclosed	March 15, 2022
GPT-4	Also trained with both text prediction and <a href="#">RLHF</a> ; accepts <a href="#">both text and images</a> as input. Further details are not public. <sup>[33]</sup>	Undisclosed. Estimated 1.7 trillion <sup>[39]</sup>	Undisclosed	March 14, 2023



# ChatGPT Examples

- ▶ GPT is capable of text understanding and cognitive reasoning
- ▶ It can answer mathematical, scientific, and literature-based questions
- ▶ Passed the BAR Exam (90th%), scored 5 on 10 AP tests, and performed very well on the SAT (93rd% on reading/writing, 89th% math)



What is the diameter of an apple divided by the thickness of an apple peel?

The diameter of an apple can vary depending on the variety, but on average it's about 7-10 cm. The thickness of an apple peel is around 1-2 mm.

Dividing the diameter of an apple by the thickness of the peel, we get:





$$7 \text{ cm} / 0.1 \text{ cm} = 70$$

So, the diameter of an apple divided by the thickness of an apple peel is approximately 70.

Note: The above numbers are estimates and the exact numbers may vary depending on the apple being considered.

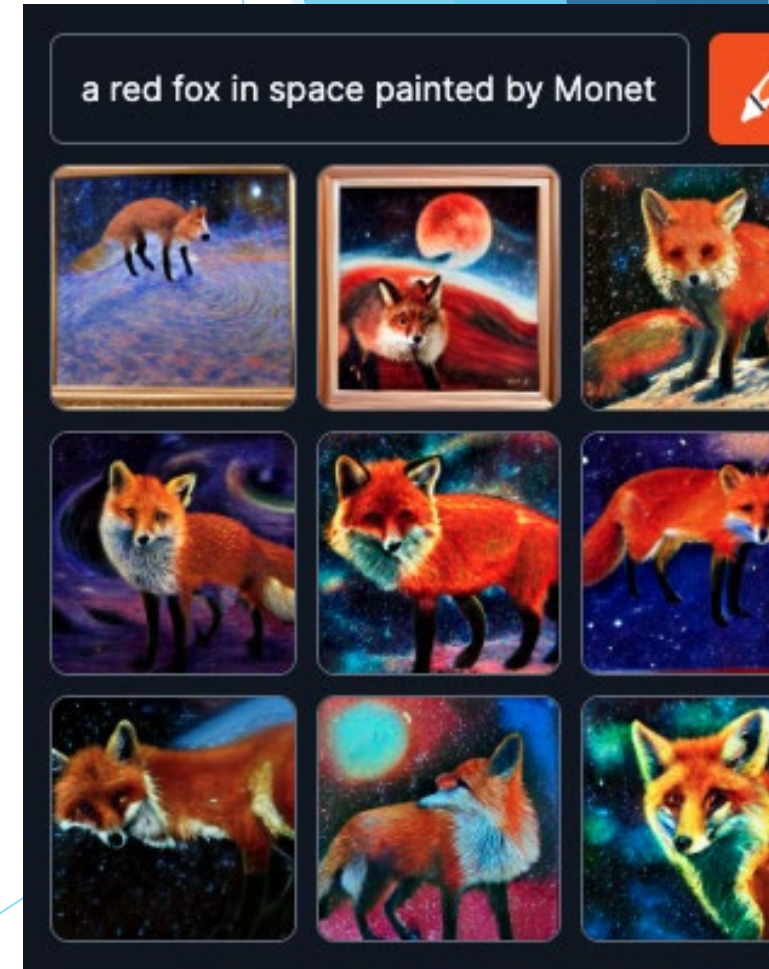
# Other GEN AI Text Models

- ▶ Facebook/Meta: Llama (70 billion Parameters)
- ▶ Google: PaLM (540 billion Parameters)
- ▶ Twitter/X: Grok (314 billion Parameters)
- ▶ Databricks: Dolly (12 billion Parameters), DBRX (132 billion Parameters)
  - ▶ DBRX was trained on 12 Trillion data points, using 3072 NVIDIA H100 GPUs

Model	GPT-3	BLOOM	LLaMA	LLaMA-2	T5	PaLM
Developer	 OpenAI	 BigScience	 Meta		 Google	
Model Size (# parameters)	175B	175B	7B, 13B, 33B, 65B	7B, 13B, 34B, 70B	11B	540B
Training Data (# tokens)	300B	350B	1.4T	2T	34B	795B

# Generative AI Applications: Image Generation

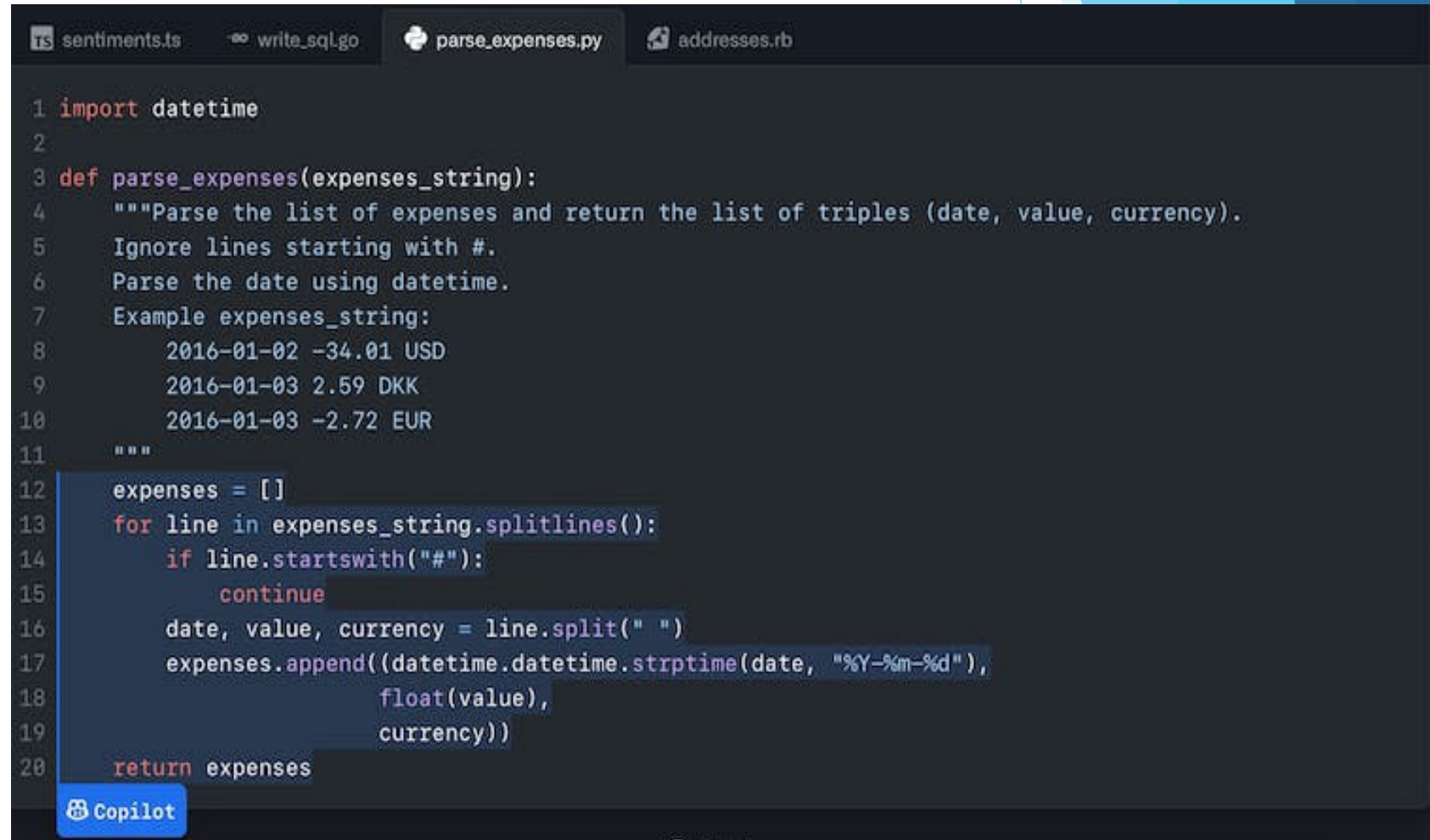
- ▶ “DALL-E” Series of models is capable of generating images from text
- ▶ DALL-E (2021): Initial model developed by OpenAI (Microsoft)
- ▶ DALL-E 2 (2022): “can combine concepts, attributes, and styles”
- ▶ DALL-E 3 (2023): “significantly more nuance and detail”
- ▶ Trained on 400 million pairs of images with text captions scraped from the Internet. Its role is to “understand and rank” DALL-E's output by predicting which caption from a list of 32,768 captions randomly selected from the dataset (of which one was the correct answer) is most appropriate for an image.
- ▶ DALL-E 2 has 3.5 billion parameters that have been optimized through its 3 day long training computation (costing ~\$300K)



# Generative AI Applications: Code Generation

## ▶ Github Copilot

- ▶ Also made by OpenAI/Microsoft
- ▶ It can generate code from English description, or complete half-done code
- ▶ Trained on 159 gigabytes of Python code sourced from 54 million public GitHub repositories



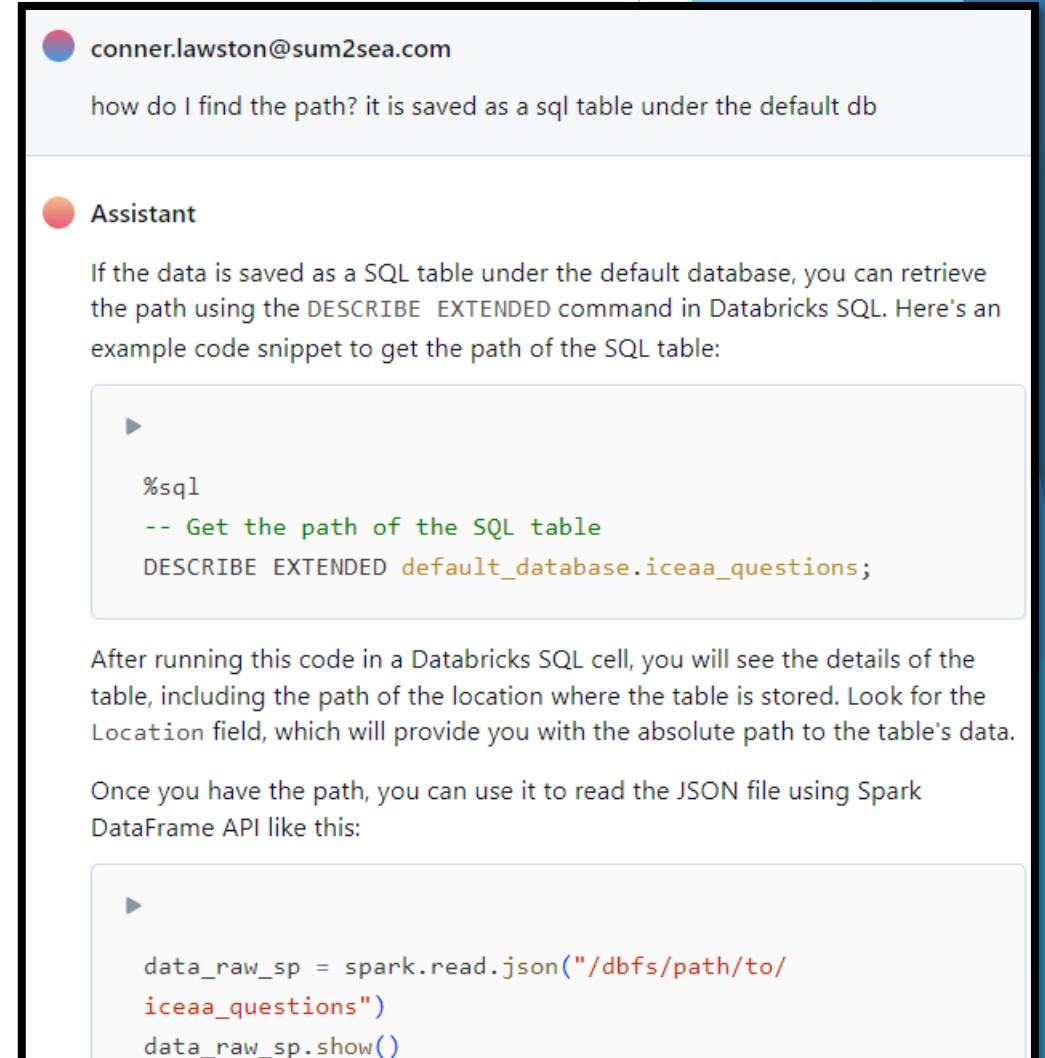
```
ts sentiments.ts  write_sql.go  parse_expenses.py  addresses.rb

1 import datetime
2
3 def parse_expenses(expenses_string):
4     """Parse the list of expenses and return the list of triples (date, value, currency).
5     Ignore lines starting with #.
6     Parse the date using datetime.
7     Example expenses_string:
8         2016-01-02 -34.01 USD
9         2016-01-03 2.59 DKK
10        2016-01-03 -2.72 EUR
11    """
12    expenses = []
13    for line in expenses_string.splitlines():
14        if line.startswith("#"):
15            continue
16        date, value, currency = line.split(" ")
17        expenses.append((datetime.datetime.strptime(date, "%Y-%m-%d"),
18                        float(value),
19                        currency))
20    return expenses

Copilot
```

# Code Generation: Databricks AI Assistant

- ▶ AI Code Generator and Debugging Assistant- Integrated into their coding environment
- ▶ It can offer fixes & improvement to your existing code
- ▶ Or just ask it questions in English, and it will generate code to perform the specified task
- ▶ Utilizes Microsoft's Azure OpenAI Service



conner.lawston@sum2sea.com

how do I find the path? it is saved as a sql table under the default db

Assistant

If the data is saved as a SQL table under the default database, you can retrieve the path using the `DESCRIBE EXTENDED` command in Databricks SQL. Here's an example code snippet to get the path of the SQL table:

```
%sql
-- Get the path of the SQL table
DESCRIBE EXTENDED default_database.iceaa_questions;
```

After running this code in a Databricks SQL cell, you will see the details of the table, including the path of the location where the table is stored. Look for the `Location` field, which will provide you with the absolute path to the table's data.

Once you have the path, you can use it to read the JSON file using Spark DataFrame API like this:

```
data_raw_sp = spark.read.json("/dbfs/path/to/iceaa_questions")
data_raw_sp.show()
```

# ChatGPT Can also be used to Generate Code

CO

Write code in base R without using any packages, to perform a iteratively reweighted least squares regression



Performing an Iteratively Reweighted Least Squares (IRLS) regression in base R without any additional packages is possible, but it's important to note that IRLS is most commonly used for generalized linear models (GLMs) with specific link functions. In this example, I'll demonstrate how to perform IRLS for a simple linear regression. Please keep in mind that for basic linear regression, the standard least squares approach is generally more appropriate.

Here's how you can perform IRLS for simple linear regression in base R:

R

Copy code

```
# Generate some sample data
set.seed(123)
x <- rnorm(100)
y <- 2 * x + rnorm(100)

# Initial values for the regression coefficients
beta0 <- 1
beta1 <- 1

# Define the model
model <- function(x, beta0, beta1) {
  beta0 + beta1 * x
}

# Perform IRLS
for (iteration in 1:100) { # Maximum number of iterations
  # Calculate the predicted values
  y_pred <- model(x, beta0, beta1)

  # Calculate the residuals
  residuals <- y - y_pred

  # Calculate the weights (inverse of the squared residuals)
  weights <- 1 / (residuals^2)

  # Update the regression coefficients
  beta1_new <- sum(weights * x * residuals) / sum(weights * x^2)
  beta0_new <- mean(weights * residuals)

  # Check for convergence
  if (abs(beta0_new - beta0) < 1e-6 && abs(beta1_new - beta1) < 1e-6) {
    break
  }

  beta0 <- beta0_new
  beta1 <- beta1_new
}

# Estimated coefficients
cat("Intercept (beta0):", beta0, "\n")
cat("Slope (beta1):", beta1, "\n")
```



# Case Study #1: Can GEN AI Models pass the ICEAA Exam?

- ▶ Certified Cost Estimator/Analyst Exam (CCEA) consists of 320 multiple choice questions
- ▶ Topics include
  - ▶ Probability & Statistics
  - ▶ Regression Analysis
  - ▶ Learning Curve Calculations
  - ▶ Inflation Indices
  - ▶ Economic Analysis
  - ▶ Earned Value Management (EVM)

# ICEAA Exam Question Example #1:

- ▶ ChatGPT Answered the question correctly and provided an explanation!



**You**

Given the hypothesis "Weight is a significant cost driver at a significance level of 0.05," which of the following statistics would you use to test for this?

- A) R-squared = 0.867
- B) CV = 15
- C) P-value = 0.022
- D) All of the above
- E) A and B only
- F) A and C only
- G) B and C only



**ChatGPT**

C) P-value = 0.022

In hypothesis testing, the p-value is used to determine the significance of a particular variable or parameter in a regression model. The p-value represents the probability of obtaining the observed results (or more extreme results) if the null hypothesis is true. In this case, the null hypothesis would be that weight is not a significant cost driver.

A significance level of 0.05 indicates that the researcher is willing to accept a 5% chance of incorrectly rejecting the null hypothesis (i.e., a Type I error). If the p-value is less than the significance level (0.05), it suggests that the observed relationship between weight and cost is statistically significant, and the null hypothesis can be rejected in favor of the alternative hypothesis (i.e., weight is a significant cost driver).

Therefore, the p-value (option C) would be used to test whether weight is a significant cost driver at a significance level of 0.05.

# ICEAA Exam Question Example #2

- ▶ ChatGPT answered this question correctly as well!



**You**

The standard for an item is 800 touch labor hours, and the documented variance factor is 18%. The corresponding indirect labor factor is 60%. What are the estimated touch labor hours for this item?"



**ChatGPT**

To calculate the estimated touch labor hours for the item, we first need to apply the documented variance factor and indirect labor factor to the standard touch labor hours.

Given:

- Standard touch labor hours = 800 hours
- Documented variance factor = 18%
- Indirect labor factor = 60%

Step 1: Adjust for the documented variance factor.

The documented variance factor is 18%, which means there's an 18% deviation from the standard.


$$\begin{aligned}\text{Adjusted touch labor hours} &= \text{Standard touch labor hours} * (1 + \text{Documented variance factor}) \\ &= 800 * (1 + 0.18) \\ &= 800 * 1.18 \\ &= 944 \text{ hours}\end{aligned}$$

# ICEAA Exam Question Example #3

- ▶ ChatGPT answered the question correctly!
- ▶ At this point, I began to worry about my job security

 **You**

You must develop a cost estimate for the structure subsystem of a new communications satellite, COM2. The new satellite is similar to COM1, a communications satellite your company recently built. Through discussion with the structural engineers, you collected the data to the right. What should the estimated cost of the structure subsystem be? COM1 Weighs 200kg and costs \$5 Million. COM2 will weigh 250kg

 **ChatGPT**

To estimate the cost of the structure subsystem for COM2, we can use the analogy estimating technique. This technique involves using historical data from a similar project (COM1) and adjusting it based on the differences between the projects (e.g., weight difference between COM1 and COM2).

Given:

- COM1 weighs 200 kg and costs \$5 million.

We need to adjust the cost based on the weight difference between COM1 and COM2.

Weight ratio = Weight of COM2 / Weight of COM1

Weight ratio = 250 kg / 200 kg = 1.25

Assuming the cost scales linearly with weight:

Estimated cost of COM2 = Cost of COM1 \* Weight ratio


Estimated cost of COM2 = \$5 million \* 1.25


Estimated cost of COM2 = \$6.25 million

Therefore, the estimated cost of the structure subsystem for COM2 is \$6.25 million.

# However, it's not Always Correct!

- ▶ On this Question, it answered wrong (even though it appears confident)
- ▶ However, the results were interesting:
  - ▶ I believe the phrase “in \$K” confused it
  - ▶ It calculated  $26.635 \times 36 = 958.86\$k$ , then added  $\$105.15 = \$959.86k$

 **You**  
If a CER for Site Development was developed giving the relationship,  $y$  (in \$K) =  $26.635x + 105.16$  (where  $x$  is the number of workstations). The predicted cost for a site that had 36 workstations would be:

 **ChatGPT**  
To find the predicted cost for a site that had 36 workstations using the Cost Estimating Relationship (CER), we substitute the value of  $x = 36$  into the equation:


$$y = 26.635 \times 36 + 105.16$$


$$y = 959.16$$

Therefore, the predicted cost for a site that had 36 workstations would be approximately \$959.16K.

So, the correct answer is:

C) \$958.86

 **You**  
Are you sure about that?

 **ChatGPT**  
Apologies for the oversight. Let's correct the calculation:

$$y = 26.635 \times 36 + 105.16$$

$$y = 959.16 + 105.16$$

$$y = 1064.32$$

Therefore, the predicted cost for a site that had 36 workstations would be approximately \$1064.32K.

So, the correct answer is:

A) \$1,064.02

# Comparing Models

- ▶ The same exam question was asked to ChatGPT and Meta's Llama-3
- ▶ However, the gave different answers as well as voice/tone
- ▶ Meta was correct... for some reason ChatGPT forgot to add the intercept

The image shows two side-by-side chat windows. The left window is for OpenAI gpt-3.5-turbo and the right is for Meta llama-3-70b-instruct-groq. Both were asked the same question: 'what would be the predicted cost for a site that had 60 workstations? Site Development:  $y$  (in K) =  $26.635x + 105.16$  (where  $x$  is the number of workstations)'. The left window lists four options (A, B, C, D) and then provides a detailed step-by-step solution, concluding with 'So the answer is A) \$1,598,100'. The right window lists the same four options and then provides a detailed step-by-step solution, concluding with '= 1703.26'.

**OpenAI gpt-3.5-turbo**

what would be the predicted cost for a site that had 60 workstations? Site Development:  $y$  (in K) =  $26.635x + 105.16$  (where  $x$  is the number of workstations)

1,598,100  
B) 1,703.26  
C) 1,742,350  
D) \$7,907.7

The predicted cost for a site with 60 workstations would be calculated as follows:

$$y = 26.635(60) + 105.16$$
$$y = 1598.1$$

Therefore, the predicted cost for a site with 60 workstations would be \$1,598,100.

So the answer is A) \$1,598,100

**Meta llama-3-70b-instruct-groq**

what would be the predicted cost for a site that had 60 workstations? Site Development:  $y$  (in K) =  $26.635x + 105.16$  (where  $x$  is the number of workstations)

1,598,100  
B) 1,703.26  
C) 1,742,350  
D) \$7,907.7

A nice linear regression problem!

We're given the equation:  $y$  (in \$K) =  $26.635x + 105.16$ , where  $x$  is the number of workstations.

We need to find the predicted cost for a site with 60 workstations. Let's plug in  $x = 60$  into the equation:

$$y = 26.635(60) + 105.16$$
$$= 1598.1 + 105.16$$
$$= 1703.26$$

# Benchmarking Methodology

- ▶ The ICEAA Exam has over 300 questions, which would take a very long time to individually copy & paste into the chat window
- ▶ Instead, it is possible to have an automated script ask the questions, record the AI's response, and score them
- ▶ The first step to this is ingesting the questions and storing them in a sql table

	Module	Question	Choices	Answer
0	Cost Estimating Basics	Which of the following are the three major types of cost estimates?	[A: Analysis of Alternatives, Budget Estimate, Independent Cost Estimate, 'B: Budget Estimate, Life Cycle Cost Estimate, Independent Cost Estimate, 'C: Independent Cost Estimate, Life Cycle Cost Estimate, Rough-Order-of-Magnitude, 'D: Budget Estimate, Independent Cost Estimate, Economic Evaluation]	B
1	Cost Estimating Basics	True or False. A Cost Analysis Requirements Description (CARD) ensures that cost projections are based on a common definition of the system and the acquisition program.	[A: True, 'B: False]	A
2	Cost Estimating Basics	A Work Breakdown Structure...	[A: Provides a financial reporting framework , 'B: Relates the work scope elements to each other and to the end product(s), 'C: Evolves as system/program evolves, 'D: All of the above]	D
3	Cost Estimating Basics	True or False. Data collection should always begin after determining the estimating methodology.	[A: True, 'B: False]	B
4	Cost Estimating Basics	Cost estimating and analysis at its best is an interdisciplinary mix which includes which of the following?	[A: Economics, Sociology, and Mathematics, 'B: Statistics, Accounting, and Political Science, 'C: Probability, Marketing, and History, 'D: Engineering, Operations Research, and Statistics]	D
...	...	...	...	...
315	Cost Management	True or False. Activity Based Costing (ABC) measures cost and performance of activities and assigns the costs of those activities to other cost objects.	[A: True, 'B: False]	A
316	Cost Management	Which of the following is true about Activity Based Costing (ABC)?	[A: Does not expose overhead costs, 'B: Relates performance to activities, 'C: Is a method for improving the visibility and causes of costs, 'D: Provides organizations with the best-suited cost view, regardless of management decisions]	C
317	Cost Management	Target costing is most closely analogous to which of the following?	[A: Activity Based Costing, 'B: Cost As an Independent Variable, 'C: Life Cycle Costing, 'D: Value Engineering]	B
318	Cost Management	Which of the following are associated specifically with Reducing Total Ownership Cost?	[A: Developing cost reduction initiatives, 'B: Allowing trade space for industry to meet user requirements, 'C: Establishing TOC baseline and cost reduction targets, 'D: Tracking progress by	F

# Asking the Questions to the AI Model

- ▶ The Databricks AI Assistant generated code to loop through each question, and ask them to ChatGPT via the OpenAI API
- ▶ This was done by prompting it in English (prompt shown in white at the top)

```
download a large language model then loop through a list of questions and ask them to the llm Accept Reject C
```

```
→ + # Import the necessary libraries
+ import openai
+ import time
+
+ # Set up your OpenAI API key
+ openai.api_key = 'YOUR_API_KEY'
+
+ # Define the function to ask questions to the language model
+ def ask_question(question):
+     # Generate a completion using the language model
+     completion = openai.Completion.create(
+         engine='davinci-codex',
+         prompt=question,
+         max_tokens=100,
+         n=1,
+         stop=None,
+         temperature=0.7
+     )
+
+     # Get the generated answer
+     answer = completion.choices[0].text.strip()
+     return answer
+
+ # List of questions to ask
+ questions = ['What is the capital of France?', 'Who is the president of the United States?', 'What is the square root of 64?']
+
+ # Loop through the questions and ask them to the language model
+ for question in questions:
+     answer = ask_question(question)
```



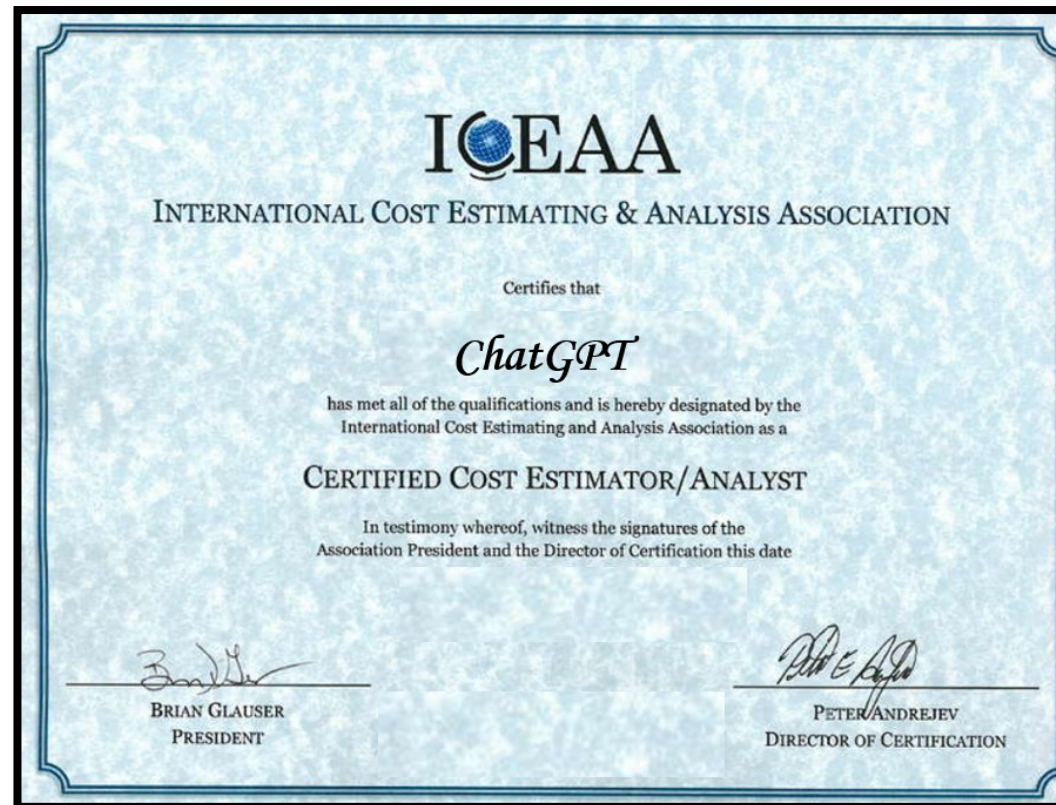
# Automated Benchmarking

- ▶ The computation time was about 8 seconds per question (perfect amount of time to get lunch and check it after)

```
-----  
Which of the following are the three major types of cost estimates?  
---  
['A: Analysis of Alternatives, Budget Estimate, Independent Cost Estimate', " 'B: Budget Estimate, Life Cycle Cost Estimate, Independent Cost Estimate", " 'C: Independent Cost Estimate, Life Cycle Cost Estimate, Rough-Order-of-Magnitude", " 'D: Budget Estimate, Independent Cost Estimate, Economic Evaluation"]  
---  
B  
-----  
True or False. A Cost Analysis Requirements Description (CARD) ensures that cost projections are based on a common definition of the system and the acquisition program.  
---  
['A: True', " 'B: False"]  
---  
A  
---  
A Work Breakdown Structure...  
---  
['A: Provides a financial reporting framework ', " 'B: Relates the work scope elements to each other and to the end product (s)", " 'C: Evolves as system/program evolves", " 'D: All of the above"]  
---  
D  
---
```

# ICEAA Exam Results

- ▶ Overall, ChatGPT was able to answer about 80% of the questions correctly!
- ▶ It should be noted that this was the generic GPT, and it had not been specifically fine-tuned for ICEAA-type questions
- ▶ In the future, it could be improved by ingesting ICEAA study materials, such as the CEBOK study guide



# Case Study #2: GURU

- LLM Specifically trained for use in Federal Contracting- developed by Summit2Sea Consulting (a cBEYONData Company)
- Input sources include 2,000 + pages of FAR & DFARS, as well as Audio Interviews with Summit2Sea Acquisition SME's
- Data is encapsulated within our cybersecurity protocols and user prompts are not kept (unlike ChatGPT)

## Introducing S2S's AI solution

**GURU**

Based on the FAR and DFAR pdf booklets, GURU will answer any questions you have about  
**Federal Regulation Acquisition**


Ask any question and see if the AI can find the correct answer!  
You can use keywords or full-fledged questions, and the AI will respond in complete sentences


 How may I help you with Federal Regulation Acquisition?

Your message



# GURU Example Output

 What is a firm fixed price contract?

 A price that is not subject to any adjustment on the basis of the contractor's cost experience in performing the contract.

Source:

Firm-fixed-price contracts. 16.202-1 Description. A firm-fixed-price contract provides for a price that is not subject to any adjustment on the basis of the contractor's cost experience in performing the contract. This contract type places upon the contractor maximum risk and full responsibility for all costs and resulting profit or loss. It provides maximum incentive for the contractor to control costs and perform effectively and imposes a minimum administrative burden upon the contracting parties. The contracting officer may use a firm-fixed-price contract in conjunction with an award-fee incentive (see 16.404) and performance or delivery incentives (see 16.402-2 and 16.402-3) when the award fee or incentive is based solely on factors other than cost. The contract type remains firm-fixed-price when used with these incentives. 16.202-2 Application.

FAR.pdf, page 476

Source is provided by GURU- where it got its answer from

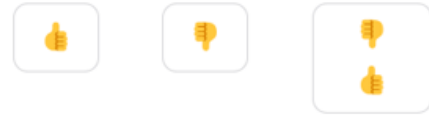
# GURU Example Output

What is fee in contracting and pricing?

39/100

Run


## Results:




Users are able to provide feedback to further improve GURU

A cost-plus-fixed-fee contract is a cost-reimbursement contract that provides for payment to the contractor of a negotiated fee that is fixed at the inception of the contract. The fixed fee does not vary with actual cost, but may be adjusted as a result of changes in the work to be performed under the contract. This contract type permits contracting for efforts that might otherwise present too great a risk to contractors, but it provides the contractor only a minimum incentive to control costs. (b) Application. (1) A cost-plus-fixed-fee contract is suitable for use when the conditions of 16.301-2 are present and, for example- (i) The contract is for the performance of research or preliminary exploration or study, and the level of effort required is unknown; or (ii) The contract is for development and test, and using a cost-plus-

# GURU Examples Continued

 What is earned value management?

 **Earned value management system means a project management tool that effectively integrates the project scope of work with cost, schedule and performance elements for optimum project planning and control.**

(iii) For cost or incentive contracts and subcontracts valued at less than \$20,000,000— (A) The application of earned value management is optional and is a risk-based decision; (B) A decision to apply earned value management shall be documented in the contract file; and (C) Follow the procedures at PGI 234.201 (1)(iii) for conducting a cost-benefit analysis. (iv) For firm-fixed-price contracts and subcontracts of any dollar value— (A) The application of earned value management is discouraged; and (B) Follow the procedures at PGI 234.201 (1)(iv) for obtaining a waiver before applying earned value management. (2) When an offeror proposes a plan for compliance with the earned value management system guidelines in ANSI/ EIA-748, follow the review procedures at PGI 234.201 (2). (3) The Defense Contract Management Agency is responsible for determining earned value management system compliance when DoD is the cognizant Federal agency.

# GURU Examples Continued



Explain the government PPBE process




Dod 7045.14

Source:

Period of Availability The period of time in which budget authority is available for original obligation. [DoD 7000.14-R Financial Management Regulation Glossary](#)

- September 2021 Planning, Programing, Budgeting, and Execution (PPBE) A cyclic process containing four distinct, but interrelated phases: planning, programing, budgeting, and execution. In addition to establishing the framework and process for decision making on future programs, the process permits prior decisions to be examined and analyzed from the viewpoint of the current environment (threat, political, economic, technological, and resources) and for the time period being addressed. The ultimate objective of PPBE is to provide operational commanders the best mix of forces, equipment, and support attainable within fiscal constraints. Political Subdivision of Local Government A local unit of government, including specifically a county, municipality, city, town, township, school district, or other special district created by State law, or combination of these units.

# GURU Examples Continued

 explain appropriations

 Appropriations used

Source:

Appropriations represent legal authority granted by Congress to incur obligations (see subparagraph 3.2.1.5 for an exception), and to make payments from the Treasury for specified purposes. An appropriations act is the most common form of budget authority; and the making of an appropriation must be expressly stated and may not be inferred or made by implication. Appropriations are available for obligation based on varying characteristics specific to periods of availability, phases of availability, and amounts. 3.2.1.1. Period of Availability. The period of time in which budget authority is available to incur new obligations. Congress establishes the period of availability. <sup>▲</sup>DoD 7000.14-R Financial Management Regulation Volume 3, Chapter 13

- February 2022 3.2.1.1.1. One-Year (Annual) Authority. Budget authority that is available for obligation only during a specified fiscal year and expires at the end of that period. For example, most Operations and Maintenance (O&M) and Personnel appropriations. 3.2.1.1.2.

Combined\_Volume1-16.pdf, page 1549



# GURU Examples Continued



What does DCMA do?



**Provides centralized, cost-effective, automated debt servicing and collection assistance to dcos on delinquent debts owed to dod by individuals who are no longer paid by dod**

Source:

Defense Contract Management Agency (DCMA) insurance/pension specialists and Defense Contract Audit Agency (DCAA) auditors assist ACOs in making these determinations, conduct CIPRs when needed, and perform other routine audits as authorized under FAR 42.705 and 52.215-2. A CIPR is a DCMA/DCAA joint review that— (1) Provides an in-depth evaluation of a contractor's— (i) Insurance programs; (ii) Pension plans; (iii) Other deferred compensation plans; and (iv) Related policies, procedures, practices, and costs; or (2) Concentrates on specific areas of the contractor's insurance programs, pension plans, or other deferred compensation plans. (b) DCMA is the DoD Executive Agent for the performance of all CIPRs. (c) DCAA is the DoD agency designated for the performance of contract audit responsibilities related to Cost Accounting Standards administration as described in FAR subparts 30.2 and 30.6 as they relate to a contractor's insurance programs, pension plans, and other deferred compensation plans. 242.7302 Requirements.

DFARS.pdf, page 717

# Conclusion on GEN AI

- ▶ Generative AI models have been in development for many years
- ▶ However, they have gained global attention and made massive strides recently
- ▶ They differ from traditional statistical models in the ways that they are built, and what output they can produce
- ▶ Textual Models can produce very confident sounding answers that are in fact incorrect (hallucinations)
- ▶ “Prompting” is an art, and can greatly improve results based on how the questions are asked
- ▶ Fine Tuning on specific data can improve results within specific topics

# Future Conversation Topics

- ▶ Will every company soon develop their own customized AI based on their proprietary data?
- ▶ Will AI models will continue to grow larger and larger with each iteration?
- ▶ If so, what will happen in terms of infrastructure and environmental repercussions?
- ▶ Where will all of this computing take place? Offshore? Undersea? In Orbit?

