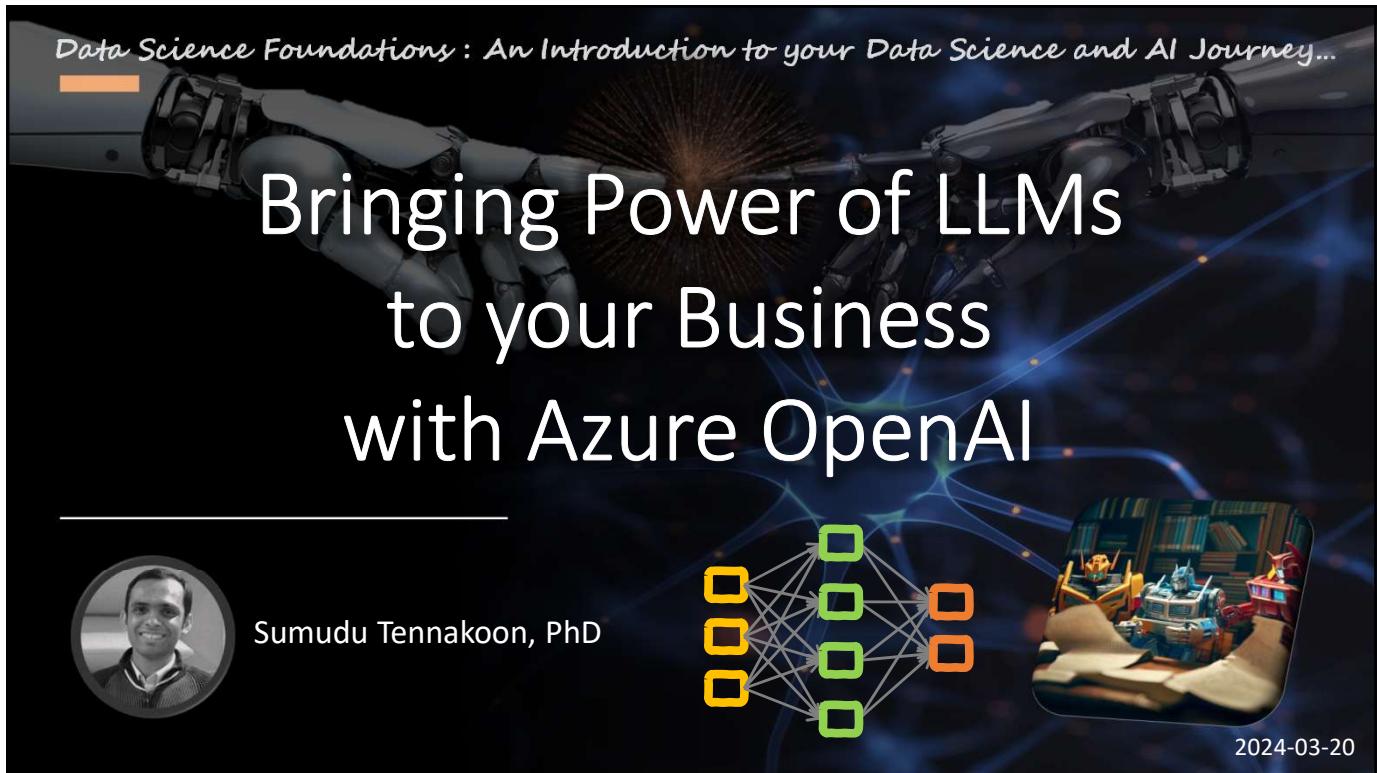


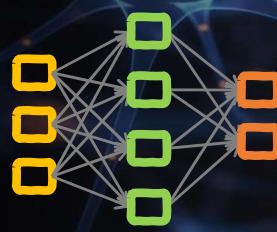
Data Science Foundations : An Introduction to your Data Science and AI Journey...



Bringing Power of LLMs to your Business with Azure OpenAI



Sumudu Tennakoon, PhD



2024-03-20

1



Outline

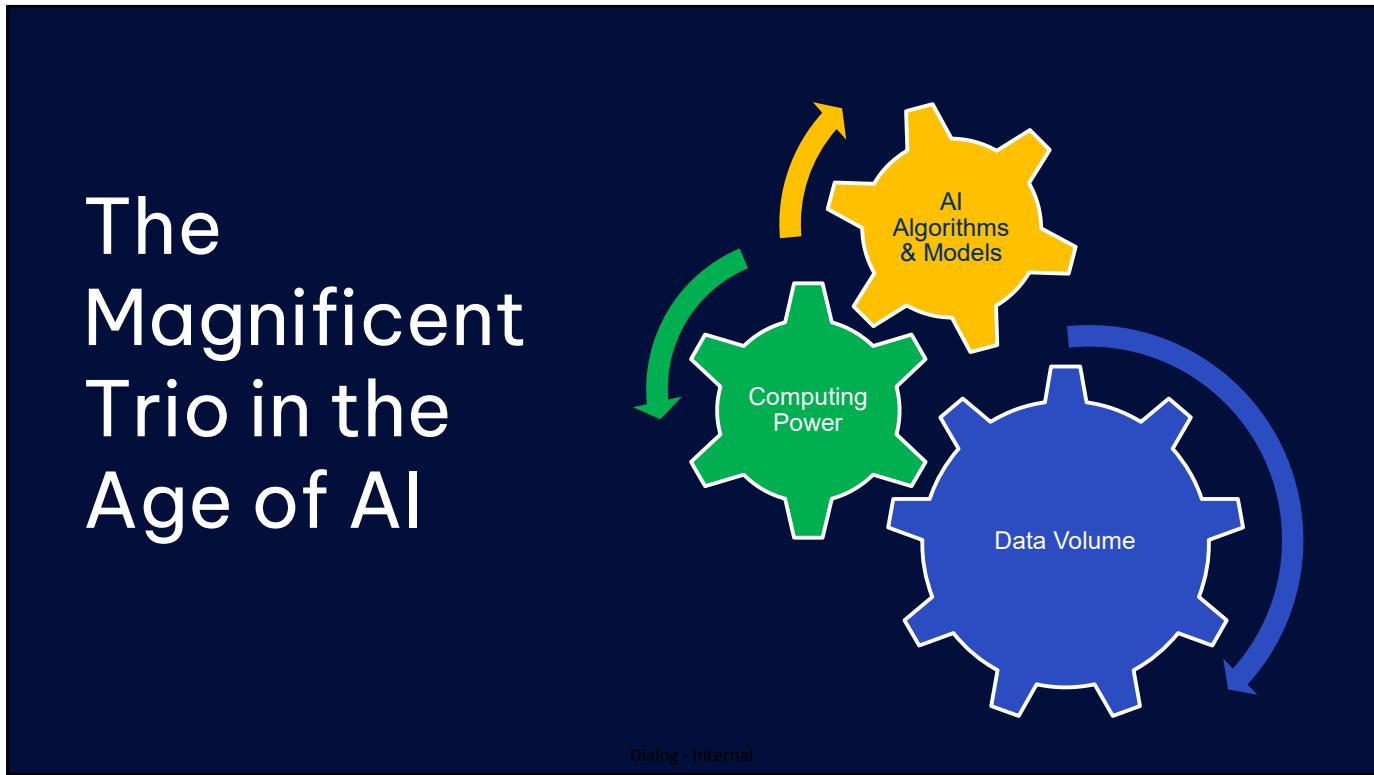
- Catching up with the Era of Copilots
- Introduction to Generative AI and Large Language Models
- Business Applications of LLMs
- Harnessing the Power of GenAI at your Business
- Getting Started with Azure OpenAI

Dialog - Internal

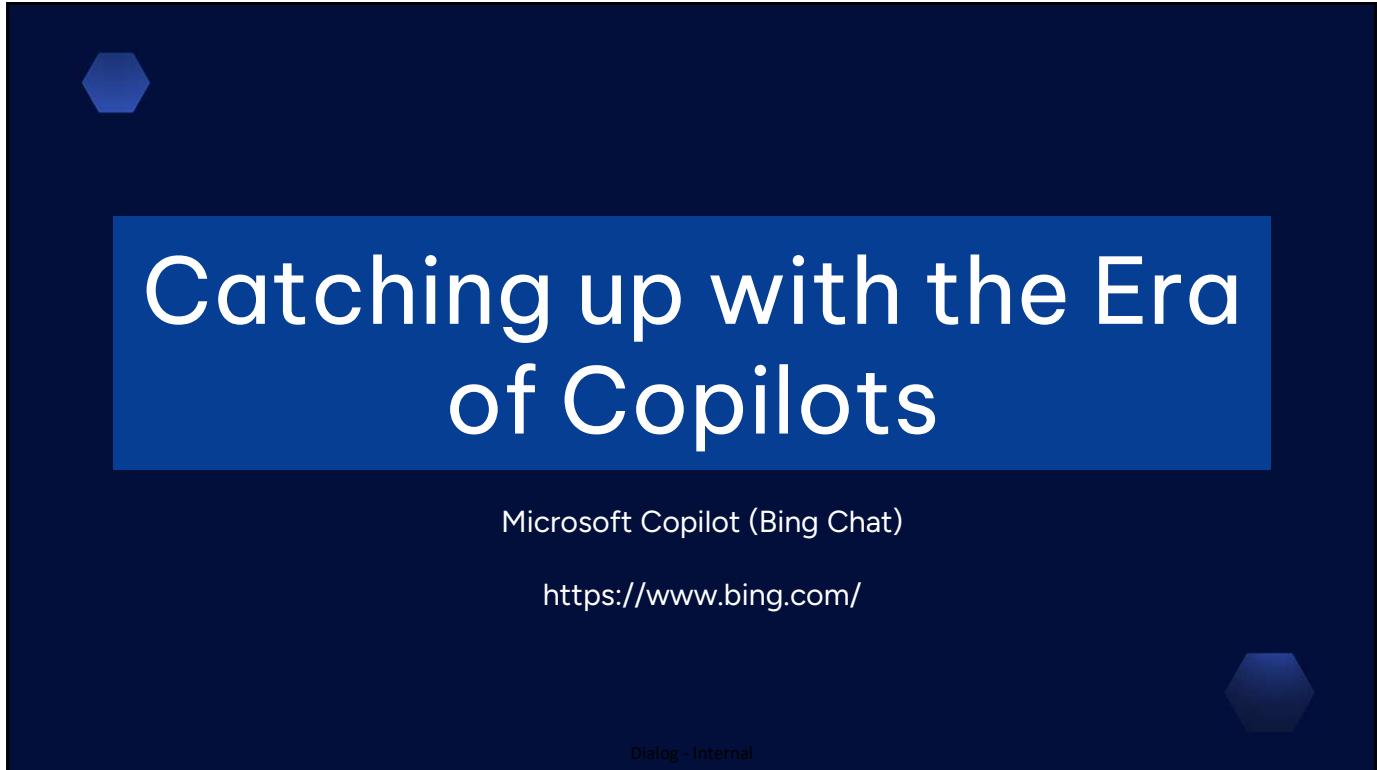
2



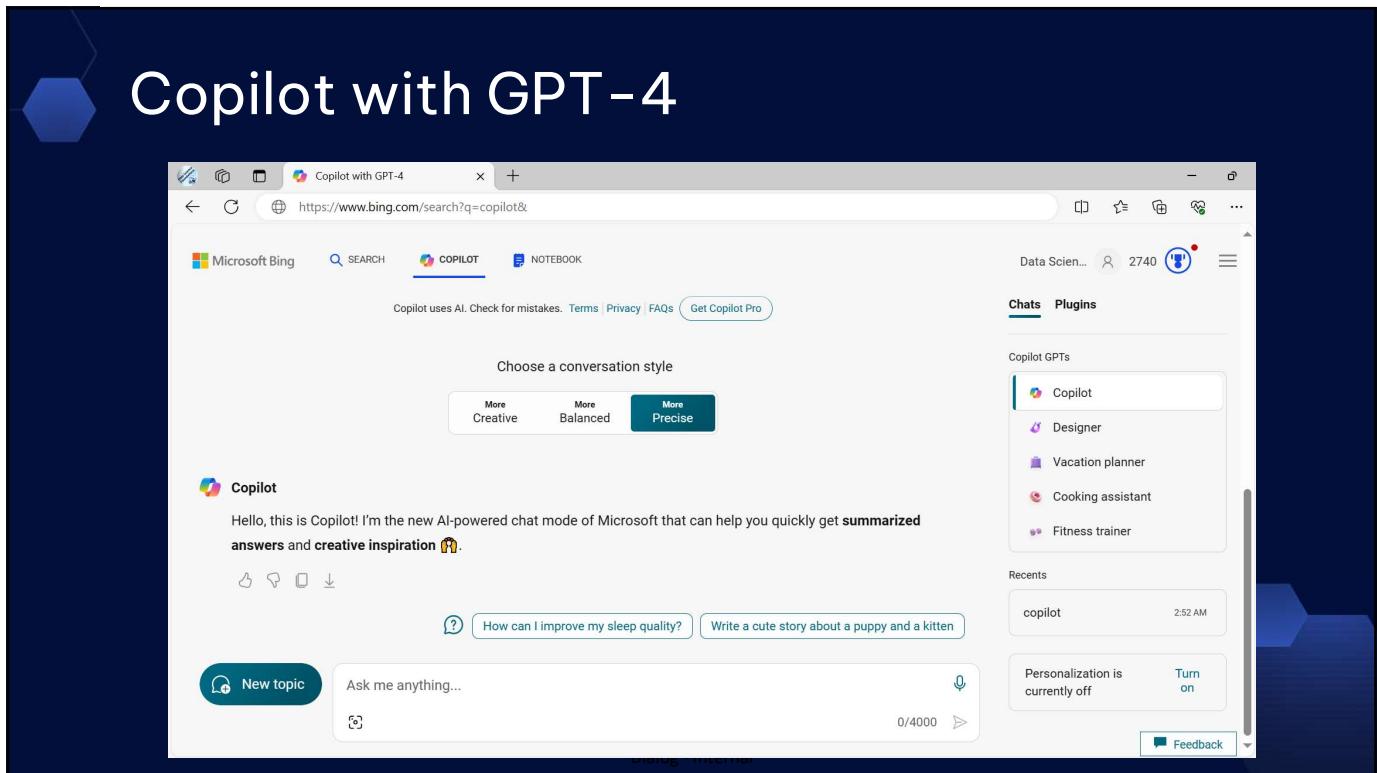
3



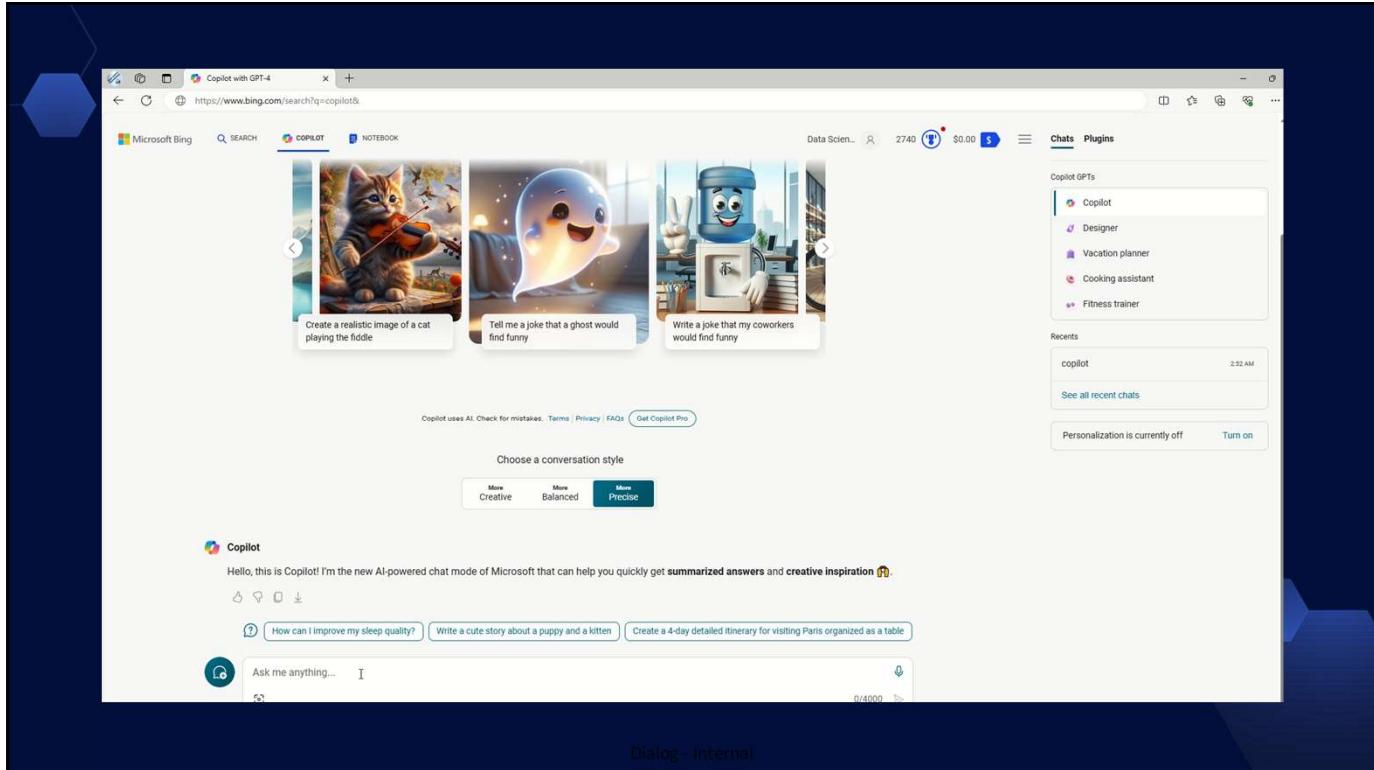
4



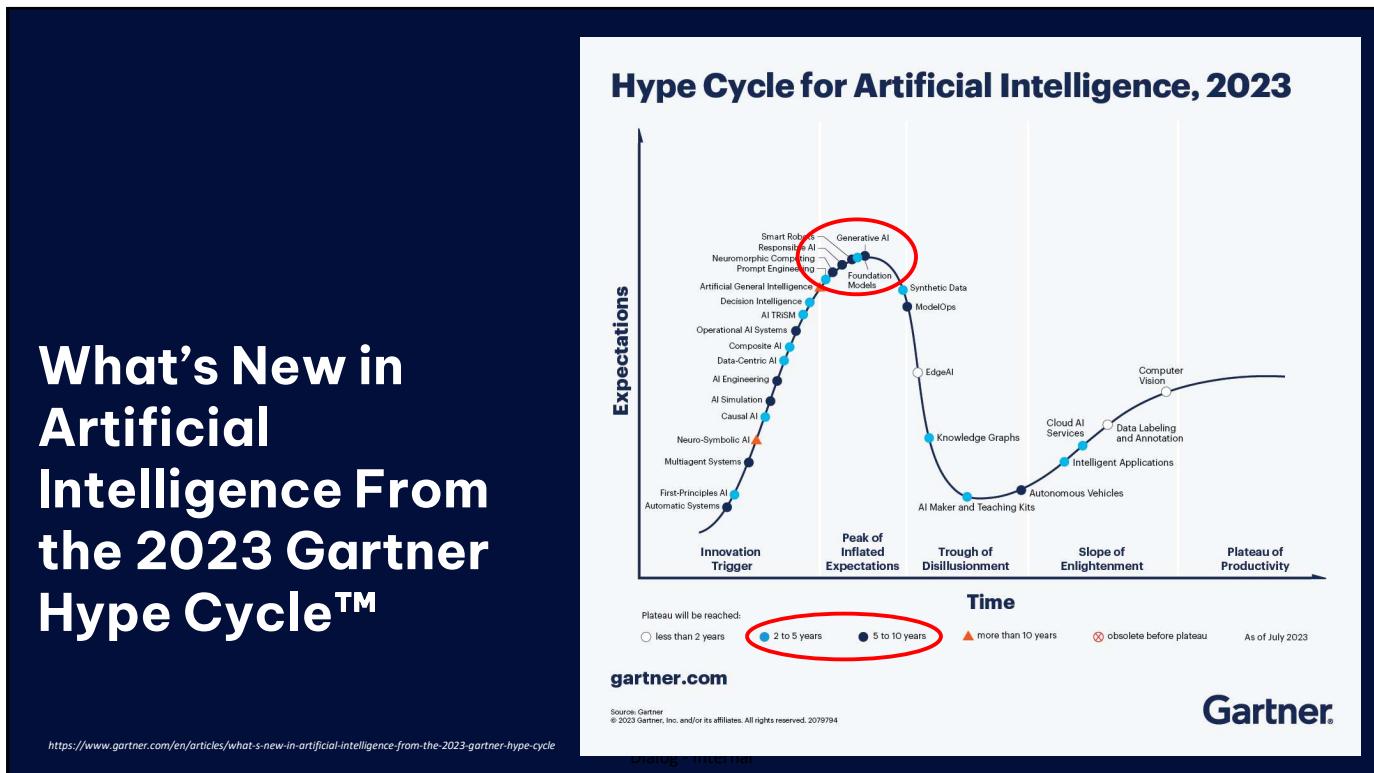
5



6



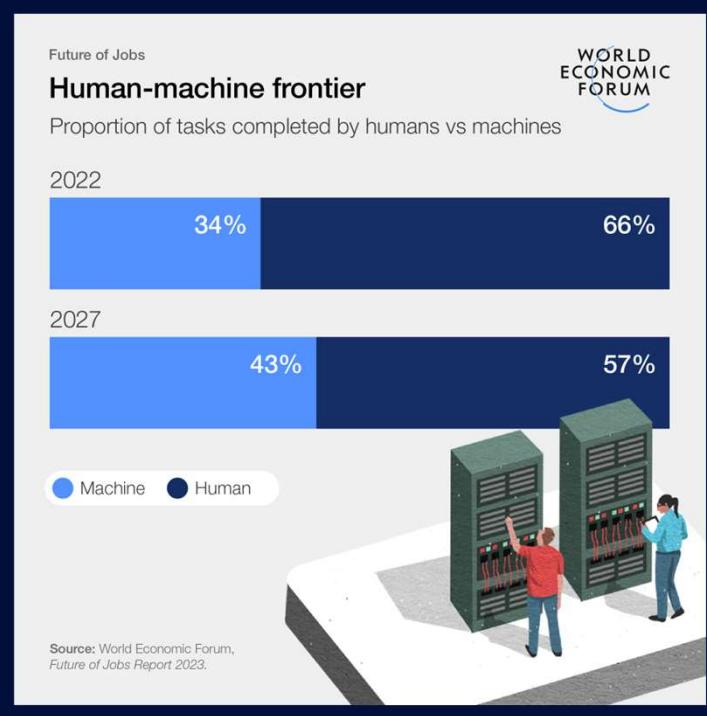
7



8

ML and AI for Automating Tasks

<https://www.weforum.org/reports/the-future-of-jobs-report-2023/infographics-2128e451e0>



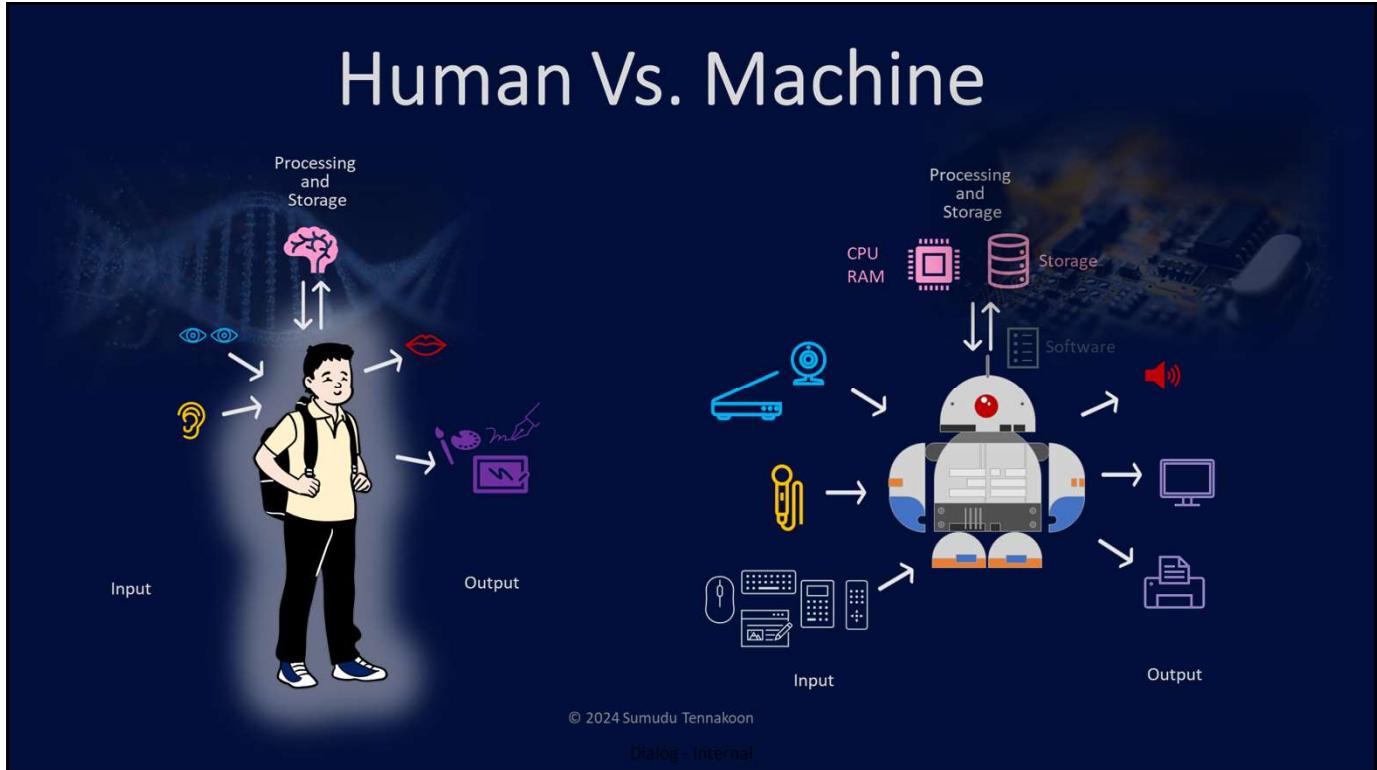
Dialog - Internal

9

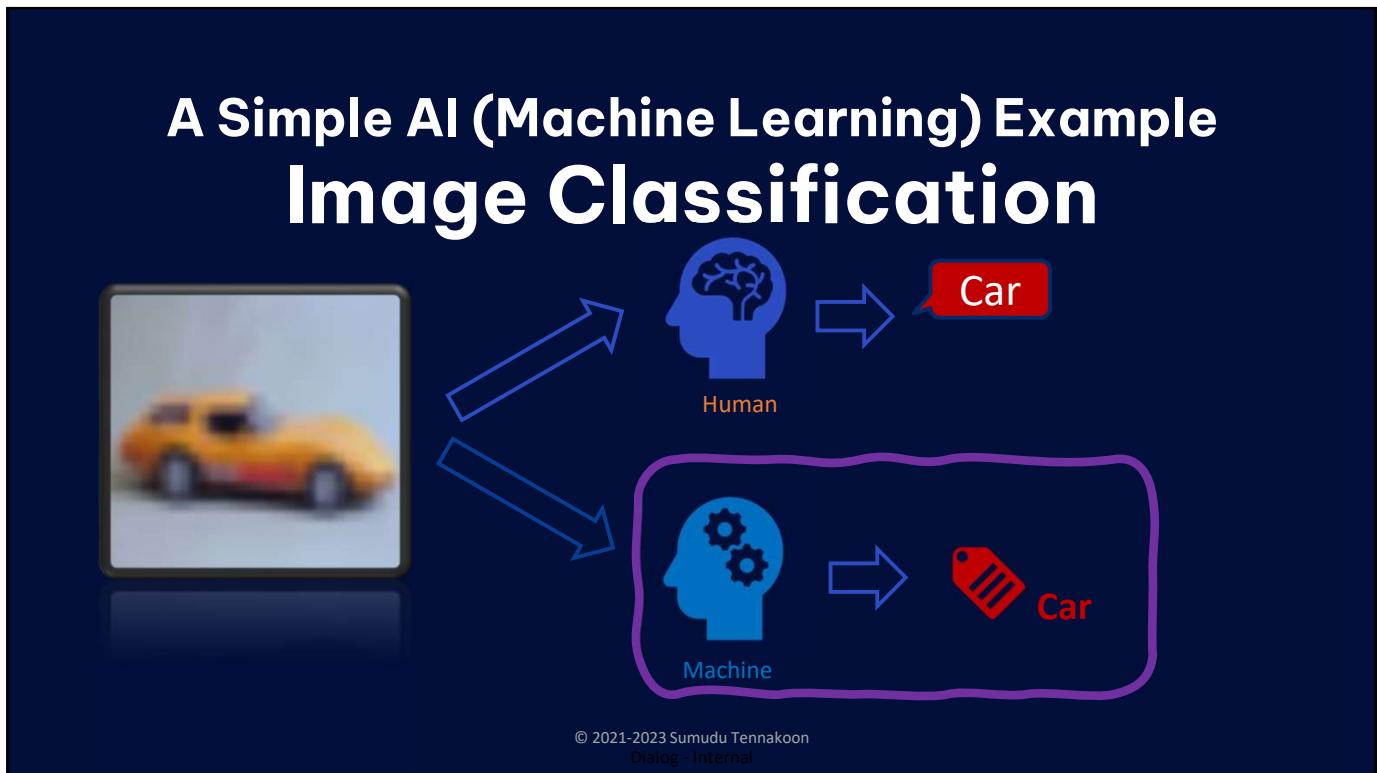
Introduction to Generative AI and Large Language Models

Dialog - Internal

10



11



12



13



14

Machine Learning vs. Generative AI

- Decision Making and Prediction
 - Classification
 - Regression
 - Clustering*
- Sequence Generation

$$y = f(X)$$

An apple fell from the <?>

$$y = f(x_{t+1}|X)$$

© 2021-2023 Sumudu Tennakoon

Dialog - Internal

word	y
sky	0.5
ground	0.1
red	0.0
tree	0.9
table	0.7

15

Generative AI

Generative AI (GAI)
is a type of AI
capable of
generating text,
images, or other
media in response
to prompts.

Dialog - Internal

16

Generative AI (GenAI)

Generative AI is a type of AI capable of generating text, images, or other media in response to **prompts**.

The diagram illustrates the hierarchy of Artificial Intelligence (AI). At the top level is "Artificial Intelligence (AI)", which contains "Machine Learning (ML)". "ML" is divided into "Traditional ML" and "Deep Learning (Neural Networks)". "Deep Learning" is highlighted with a red arrow pointing to it. Below "ML" is "Natural Language Processing". At the bottom level is "Data Science", which includes "Dialog - Internal". To the left of the main stack, there is a section titled "Sensors" with icons for eye, ear, mouth, hand, and brain.

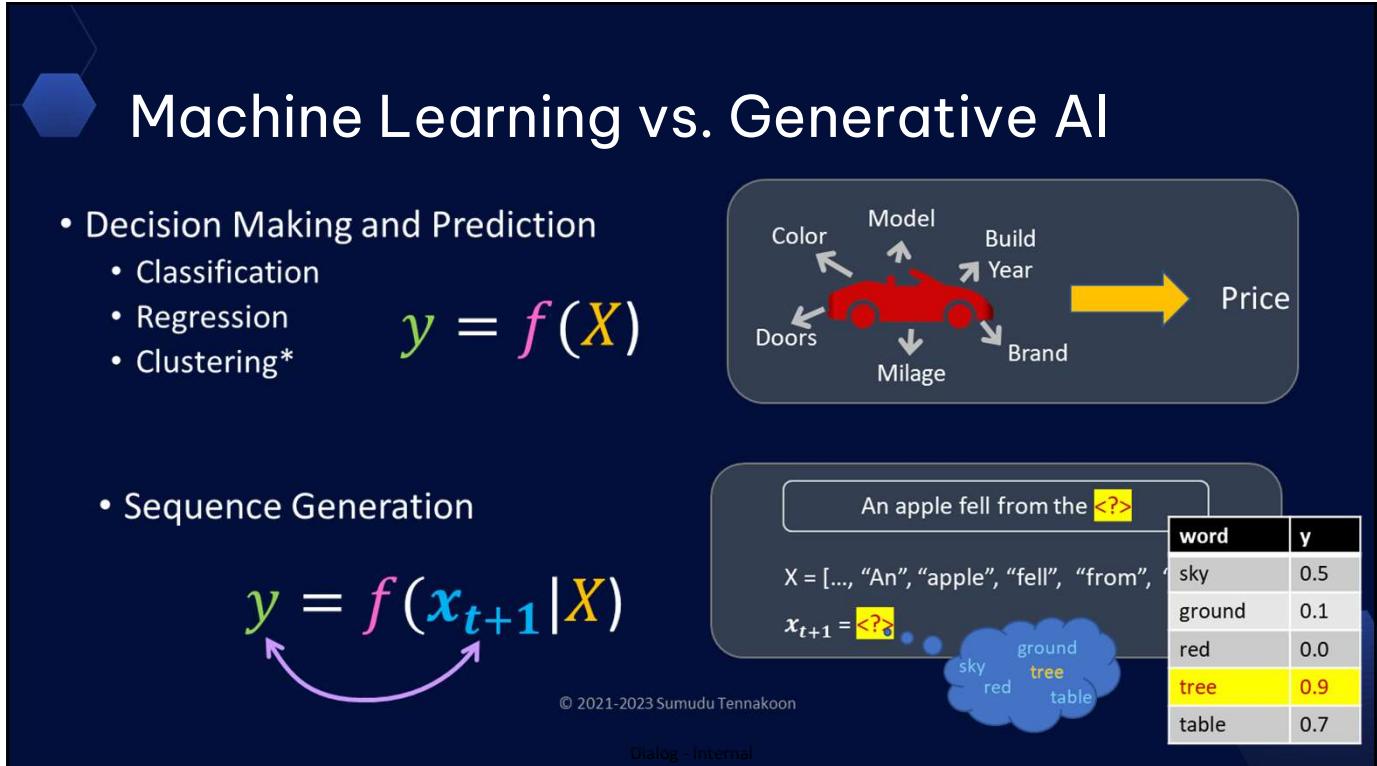
17

Generative AI Applications

The diagram shows four categories of Generative AI Applications:

- Natural Language Processing** (Icon: Document):
 - Q & A Bot
 - Summarization
 - Translation
 - Sentiment Analysis
 - Generate Response to Email
 - Information Extraction
 - Code Generation*
- Computer Vision** (Icon: Camera):
 - Image classification
 - Object Detection
 - Semantic Segmentation
 - Image Restoration
- Audio** (Icon: Sound waves):
 - Speech Recognition
 - Audio Classification
 - Music Generation
 - **Voice Generation**
- Multimodal tasks** (Icon: Jigsaw puzzle):
 - Image Captioning
 - Document Question Answering
 - Image Generation by Text Prompt
 - Video Classification

18



Machine Learning vs. Generative AI

- Decision Making and Prediction
 - Classification
 - Regression
 - Clustering*
- Sequence Generation

$y = f(X)$

Model
Build Year
Price
Color
Doors
Milage
Brand

An apple fell from the <?>

X = [..., "An", "apple", "fell", "from",
 $x_{t+1} = \text{<?>} \dots$

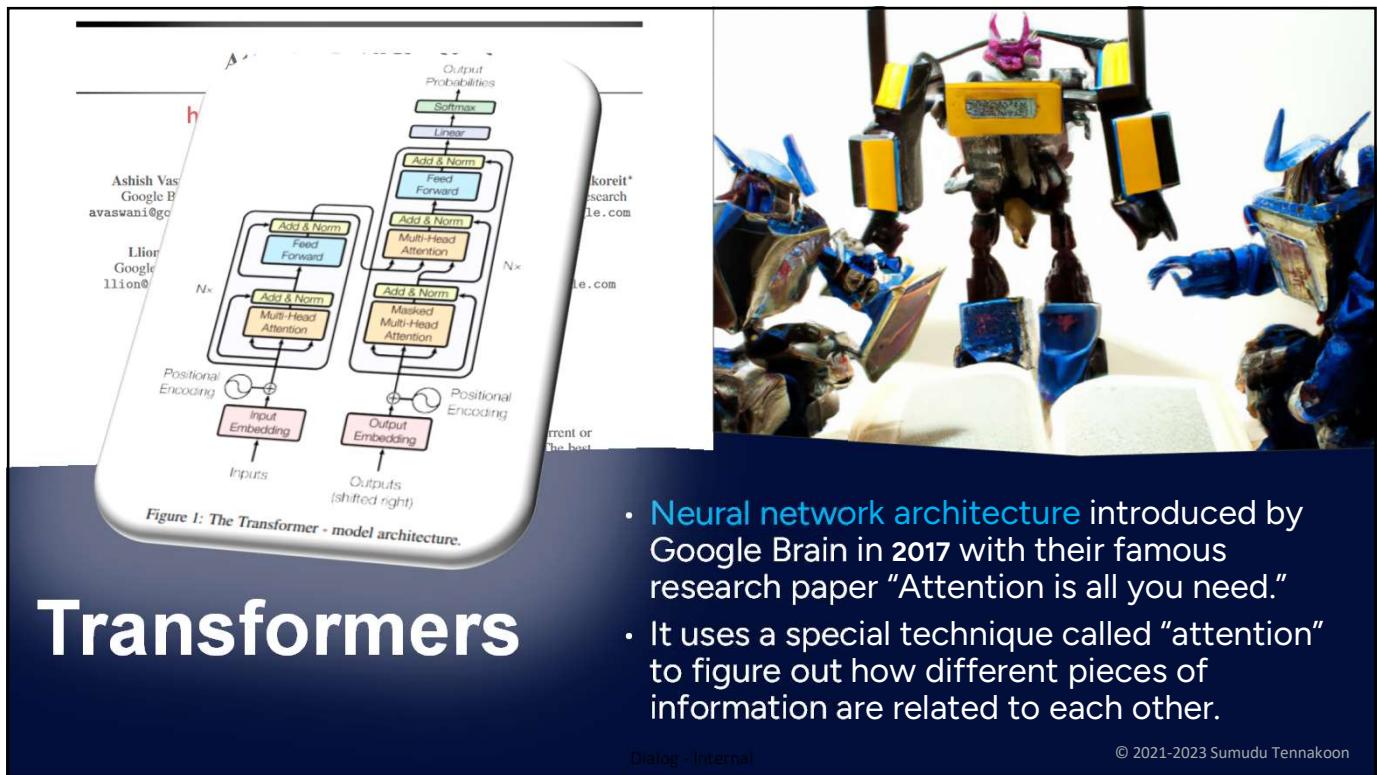
word | y

sky	0.5
ground	0.1
red	0.0
tree	0.9
table	0.7

Dialog - Internal

© 2021-2023 Sumudu Tennakoon

19



Ashish Vaswani, Google Brain, avaswani@google.com
Llion Liao, Google Brain, llion@google.com

Figure 1: The Transformer - model architecture.

The diagram illustrates the Transformer model architecture. It shows the flow of data through input embeddings, position embeddings, and multiple layers of multi-head attention and feed-forward networks, followed by a linear layer and softmax output. The architecture is labeled with N and h .

Outputs (shifted right)

Inputs

Positional Encoding

Add & Norm

Multi-Head Attention

Feed Forward

Linear

Softmax

Output Probabilities

Masked Multi-Head Attention

Add & Norm

Multi-Head Attention

Feed Forward

Linear

Softmax

Output Probabilities

Positional Encoding

Outputs (shifted right)

Figure 1: The Transformer - model architecture.

Transformers

• Neural network architecture introduced by Google Brain in 2017 with their famous research paper "Attention is all you need."

• It uses a special technique called "attention" to figure out how different pieces of information are related to each other.

Dialog - Internal

© 2021-2023 Sumudu Tennakoon

20

Foundation Models

Foundation Models are Generative Artificial Intelligence Models trained on a vast amount of unlabeled data and can be adapted to perform a broad range of tasks.

Examples of Foundation Models are Large Pretrained Language models like BERT, GPT-3, BLOOM, Bard, Llama, Mistral; Vision Models like CLIP, ALIGN and Multimodal models like GPT-4 and DALL-E.



21

Foundation Models

Foundation Model

Labeled Data

Unlabeled Data

Fine-tuning
(Optional)

Text Generation

Q & A

Information Extraction

Text Classification

Sentiment Analysis

Translation

© 2021-2023 Sumudu Tennakoon
Dialog - Internal

22

DF0

Modalities

Type of data set used to train a generative AI system

- Text (Natural Language)
- Images (Vision)
- Code
- Music/Speech (Audio)
- Video
- ...



The capability of AI system is limited by the modality.

© 2021-2023 Sumudu Tennakoon

23

Large Language Models (LLMs)

- What are they?
 - Machine learning Models
 - Process and Generate natural (human) language.
- Why we call them "Large"?
 - Made of a neural network with large number of parameters.
 - Trained on large amount of text data.
- Why we call them "Language Models"
 - Trained on text data (Text modality)
 - Capable of processing and generating natural language.
- How they got trained?
 - Huge amounts of text data, such as books, articles, social media posts, and more, to learn the patterns and structures of natural language
- What can they do?
 - Text generation, summarization, translation, question answering, and sentiment analysis.
 - Can be generalized for many applications (Foundational Models)

© 2021-2023 Sumudu Tennakoon
Dialog - Internal

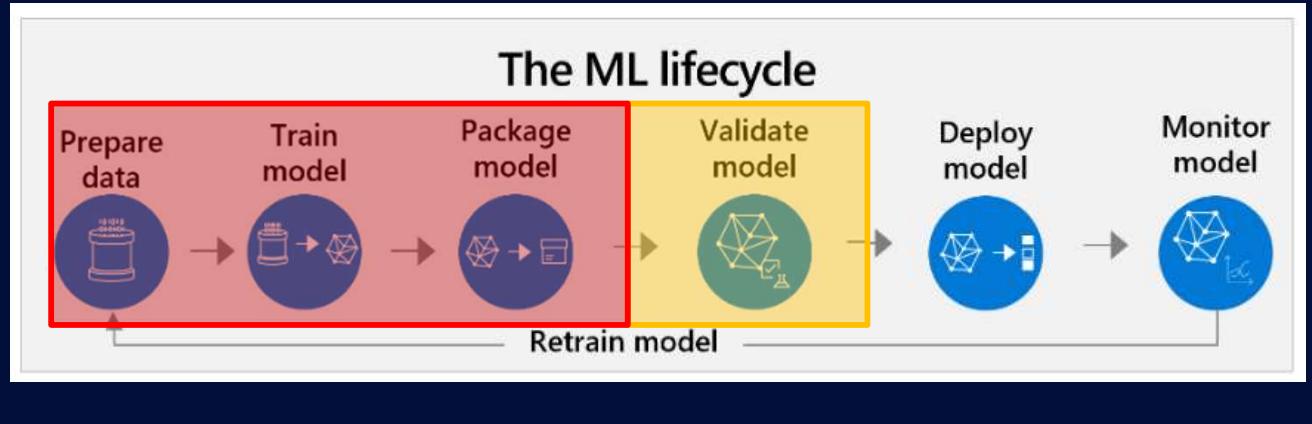
24

Slide 23

DFO Ake it 1 min

Data Science Foundations, 2023-06-04T04:05:37.540

Advantage of Using Pre-trained Models

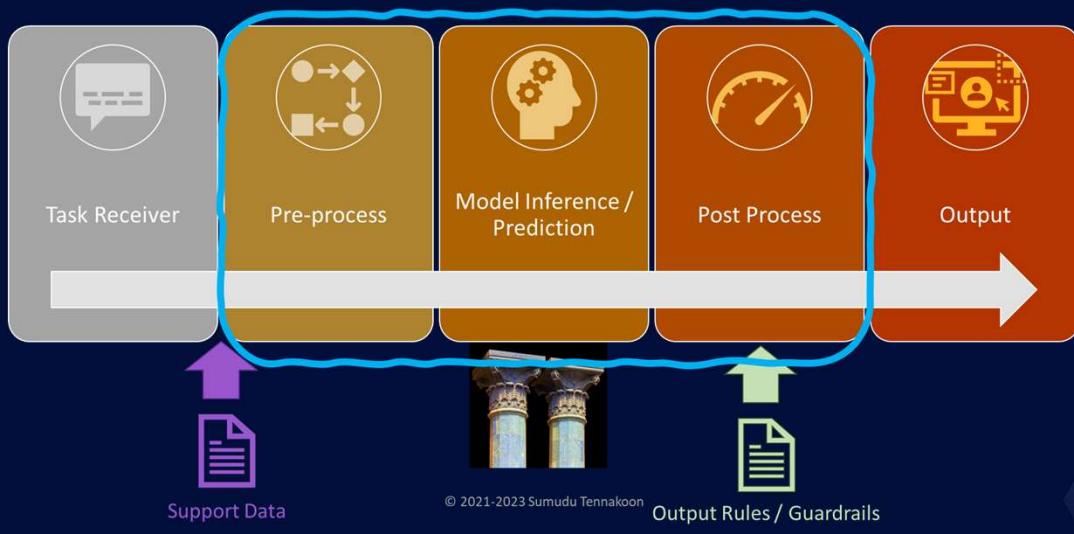


[Discover how Microsoft Power Platform brings AI to your business - Training | Microsoft Learn](#)

Dialog - Internal

25

LLM Application Pipeline

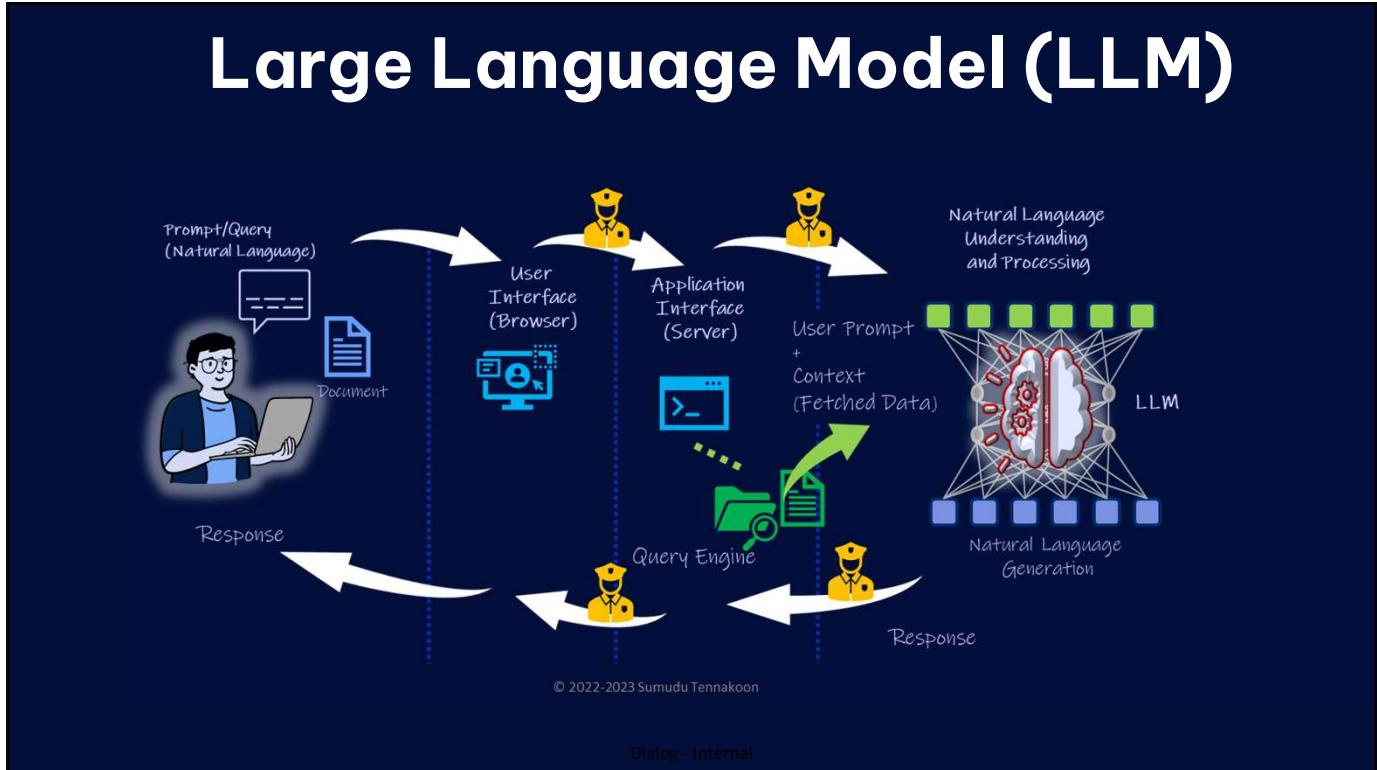


© 2021-2023 Sumudu Tennakoon

Output Rules / Guardrails

Dialog - Internal

26



27



28

Email Response Generation

Prompt

Write a reply to this email:
Dear John,

It was great doing business with you.
Please ship order to my home address
2111 Ash Lane, Crestview CA 92002.

My customer account number is
Z1234567890.

Thanks,

Maya
XYZ Company LLC
maya@xyz.com
o: +1 800 234 5678
www.xyz.com

Dear

Response

© 2021-2023 Sumudu Tennakoon
Dialog - Internal

Dear Maya,

Thank you for your business. I have received your order and your customer account number.

I will ship the order to the address you provided: **2111 Ash Lane, Crestview CA 92002**. We will use the account number **Z1234567890** to process and track your order.

Thank you again for choosing **XYZ Company LLC**. We look forward to serving you in the future.

Thank you,
John

29

Querying Data with Generative AI

Prompt

In our HR database we have the following tables. **Employee** table with **emp_id** as the primary key and other columns are **dept_id**, **hourly_rate** and **age**. **Department** Table has **dept_id** column as primary key and it also has **department_name** column. Generate a SQL query to get count of employees by department name.

Natural Language

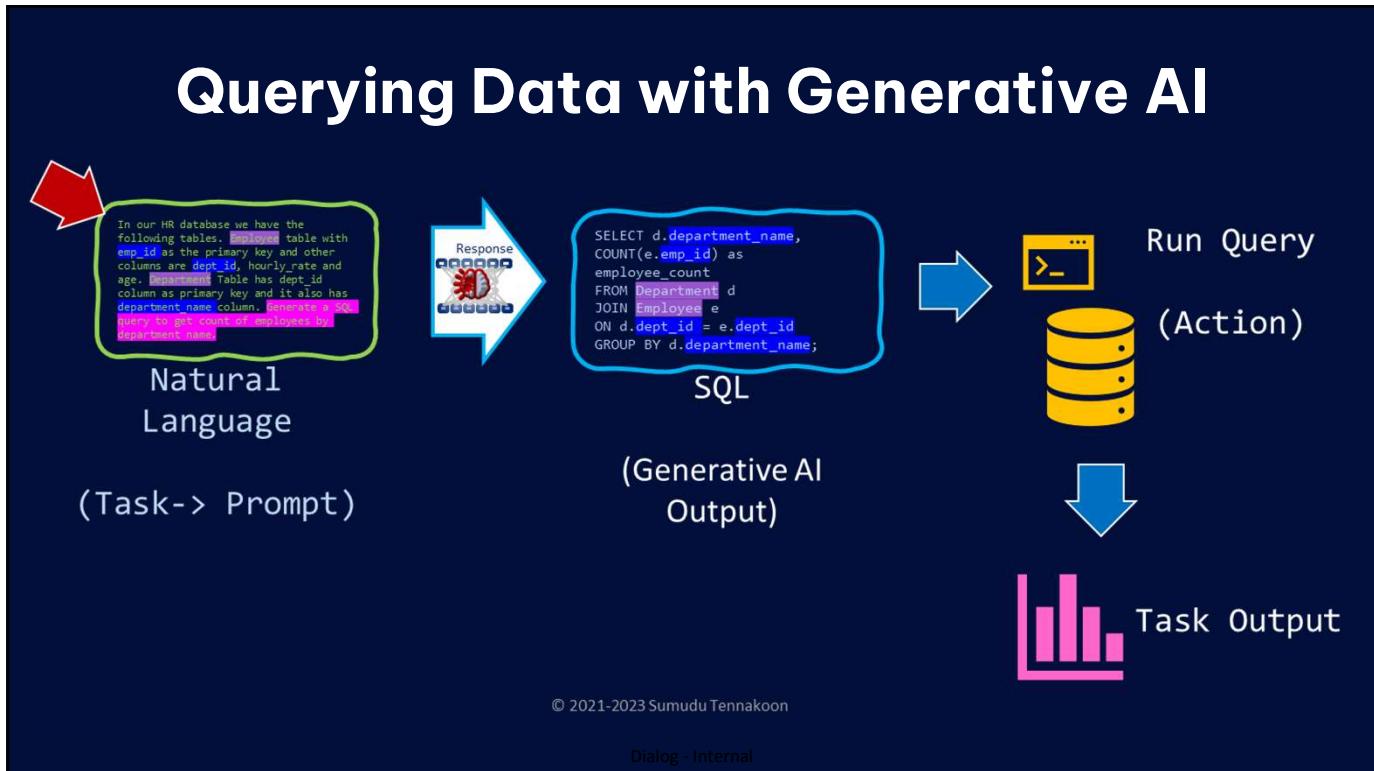
Response

SQL Query

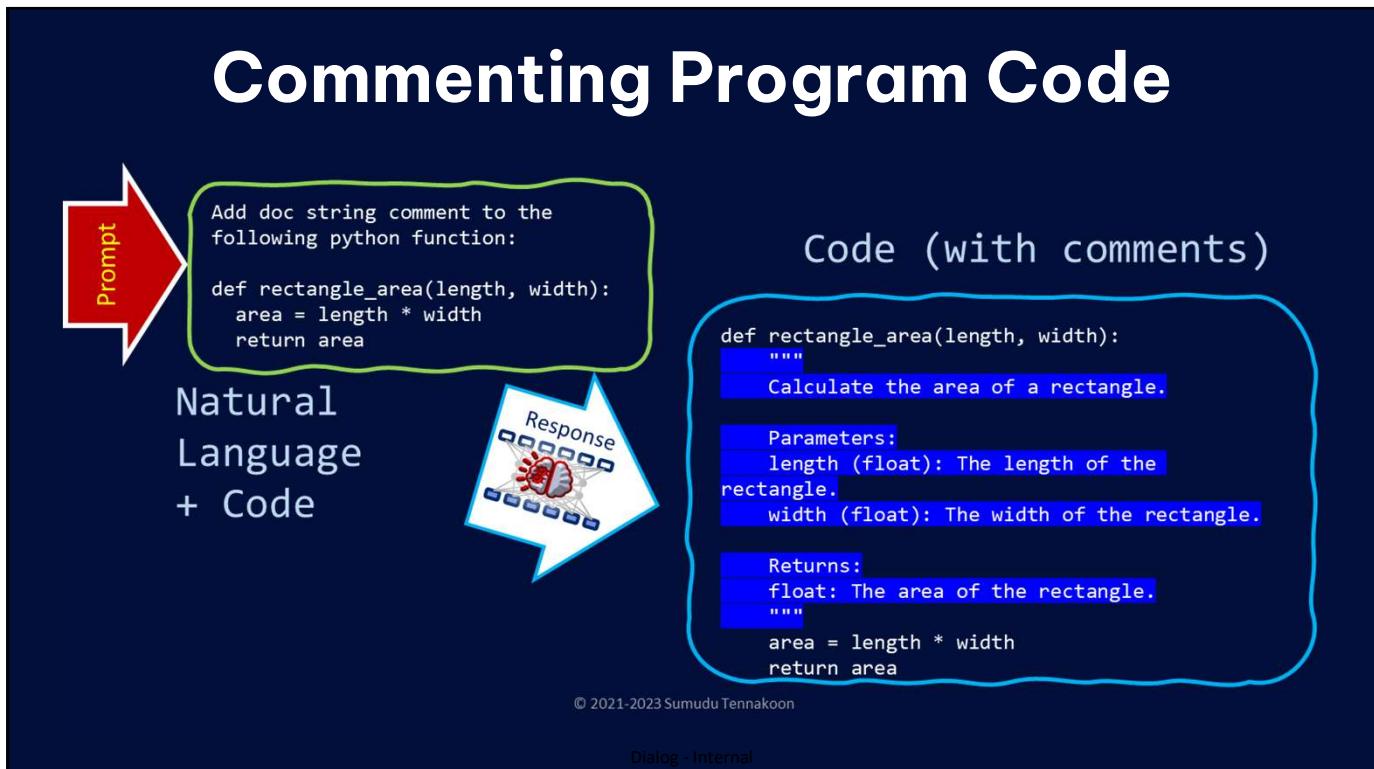
```
SELECT d.department_name, COUNT(e.emp_id) as employee_count
FROM Department d
JOIN Employee e
ON d.dept_id = e.dept_id
GROUP BY d.department_name;
```

© 2021-2023 Sumudu Tennakoon
Dialog - Internal

30

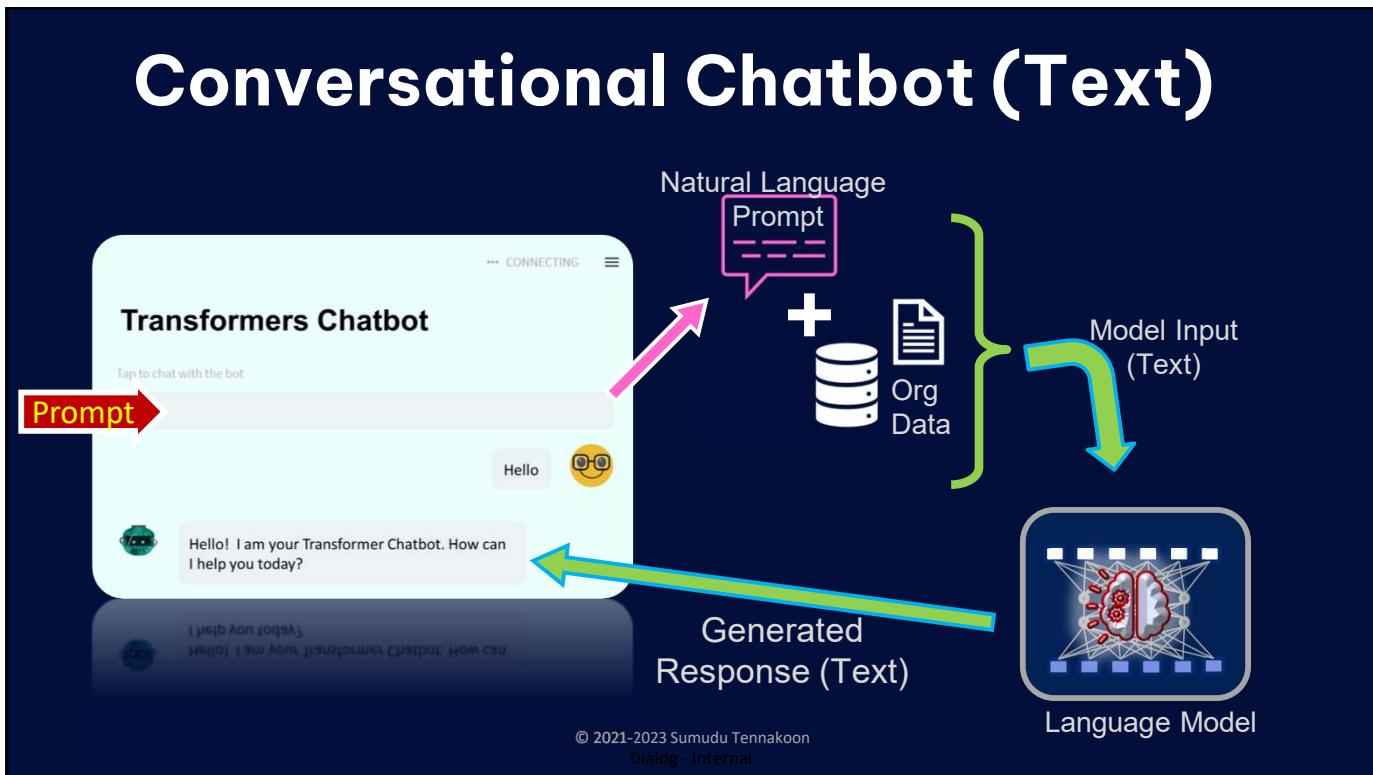


31



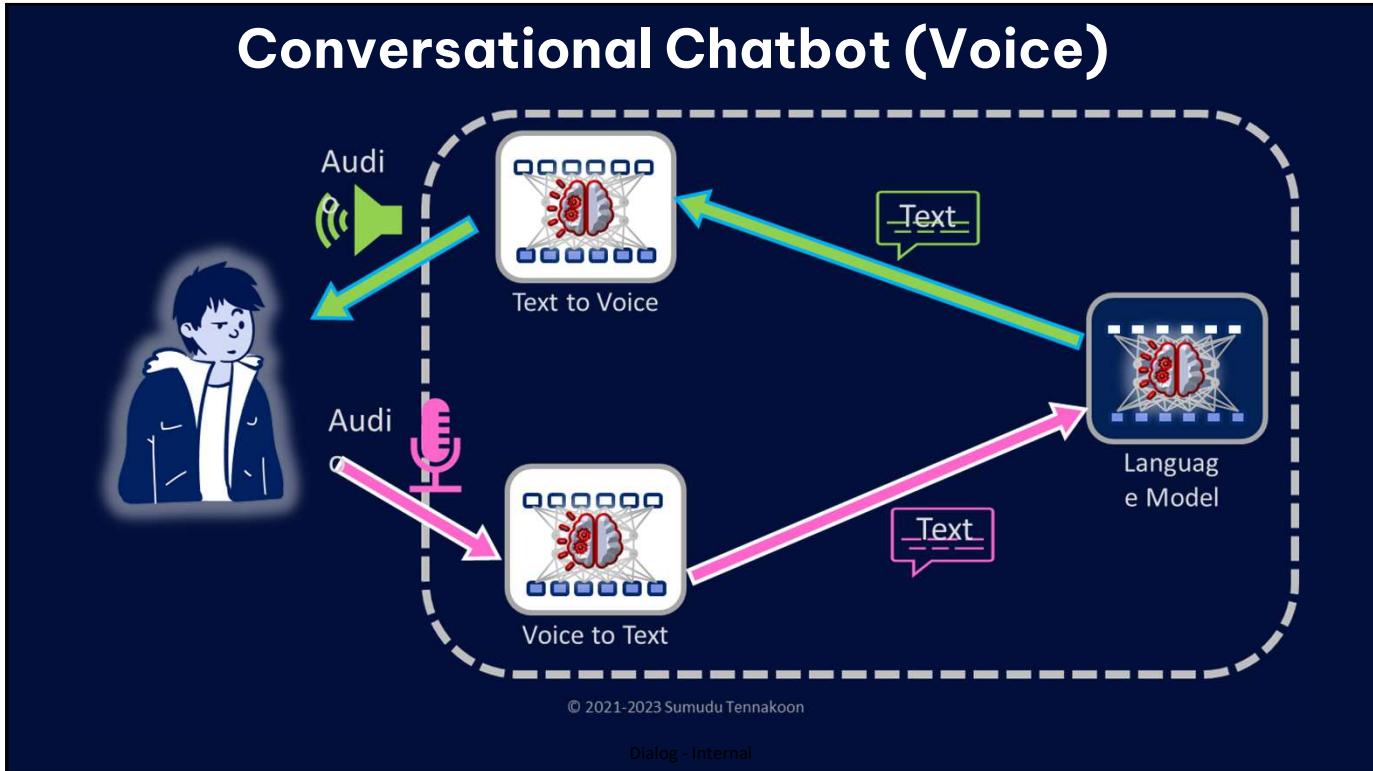
32

Conversational Chatbot (Text)

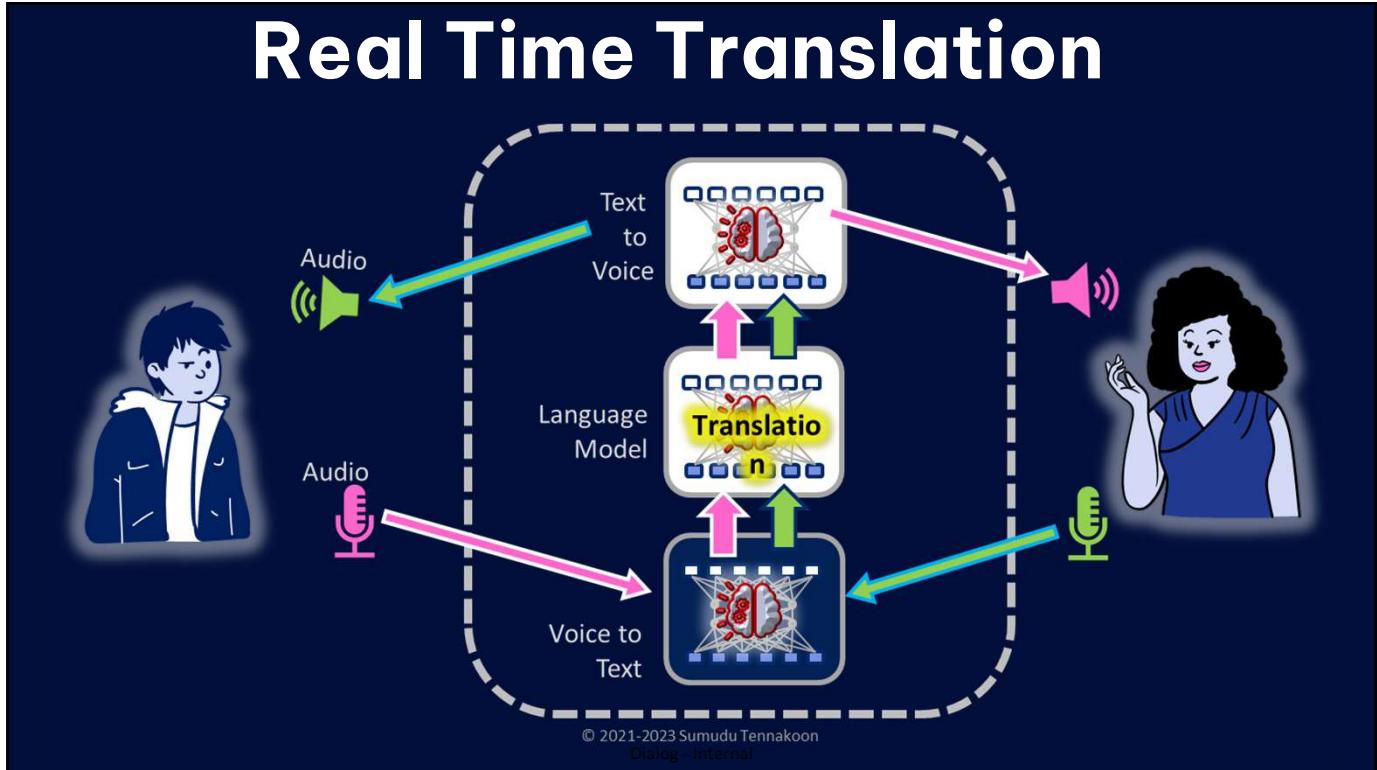


33

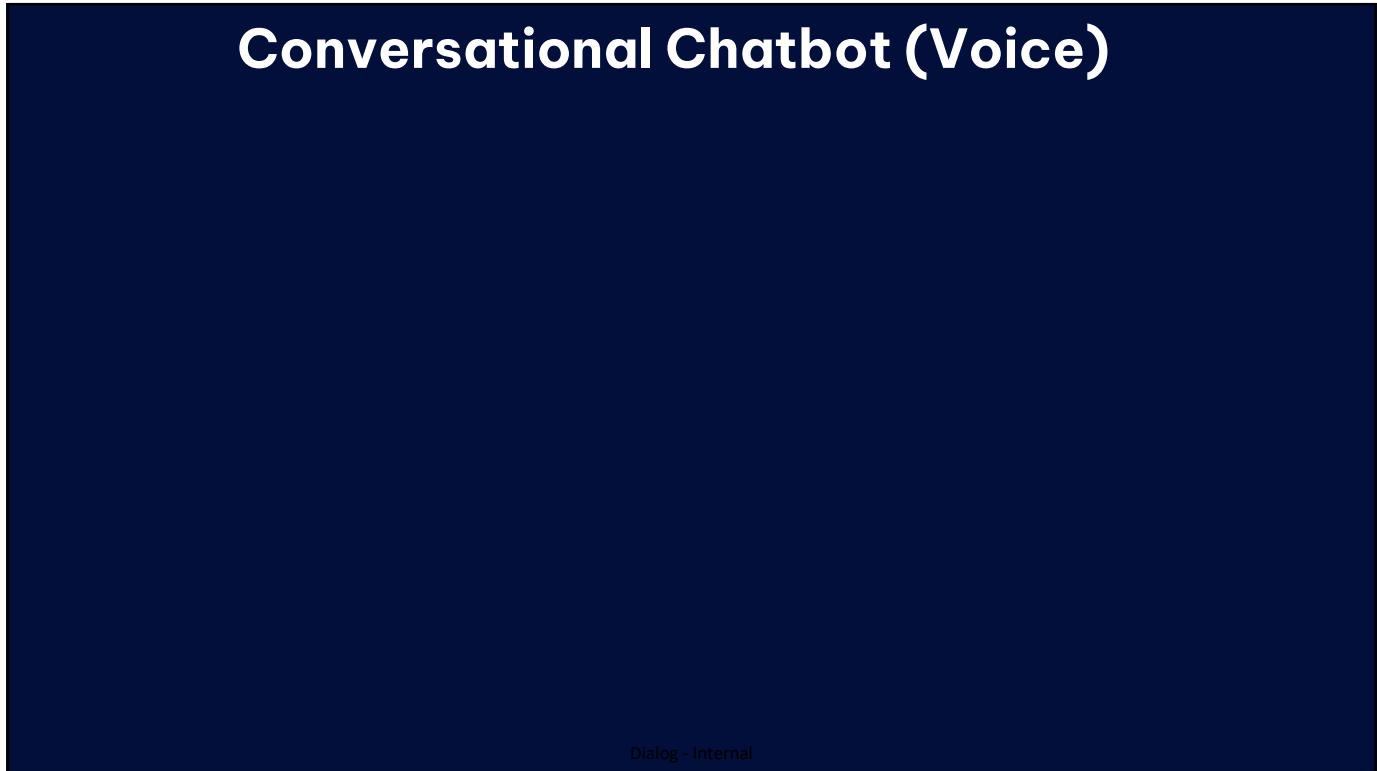
Conversational Chatbot (Voice)



34



35



36

Sentiment Analysis

A diagram illustrating sentiment analysis. On the left, there is a blue washing machine icon surrounded by six orange human icons, each with a speech bubble containing a short text snippet. To the right, four colored callout boxes show the resulting sentiment scores:

- Green box: "A good product for a reasonable price. Loved to have it my home."
- Red box: "Item broke within two weeks of use. Customer service is not helpful. I would not buy this product again."
- Yellow box: "You will get what you paid for. Works just fine."
- Light green box: "A great product. Well designed. !"

Below the diagram, the text "Dialog - Internal" and "© 2024 Sumudu Tennakoon" is visible.

37

Asking Questions from Own Documents

A diagram illustrating how to ask questions from own documents. On the left, a cartoon character asks a question, which is processed by a computer (represented by code brackets </>) and then retrieved from a "Document Collection" (represented by a stack of documents). The process involves bidirectional communication between the user, the computer, and the document collection.

Below the diagram, the text "© 2021-2023 Sumudu Tennakoon" and "Dialog - Internal" is visible.

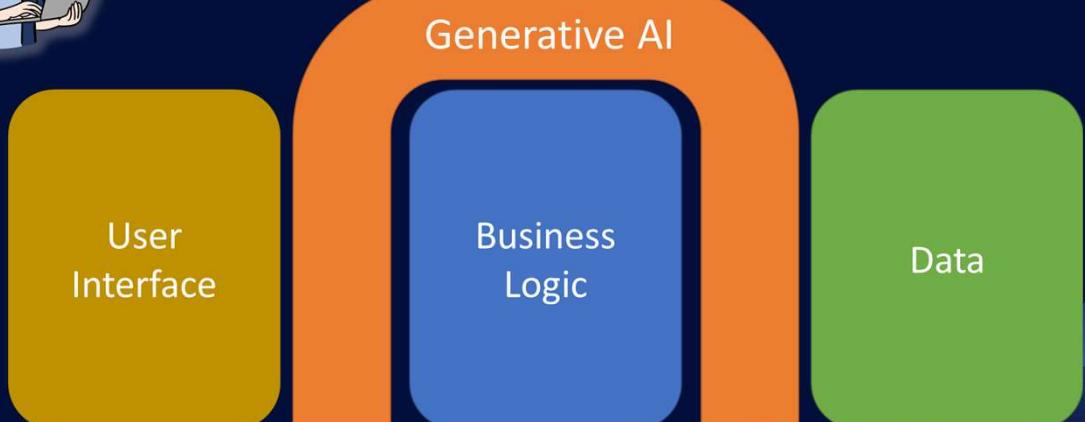
38

Harnessing the Power of GenAI at your Business

Dialog - Internal

39

Harnessing the Power of GenAI



Dialog - Internal

40

Getting Started with Generative AI

- Build vs. Buy
- Cloud AI Services
- Open-Source Frameworks
- Pre-trained Models
- Model Customizing/Fine Tuning
- Build Models From Scratch
- Integrating Business Data Layer

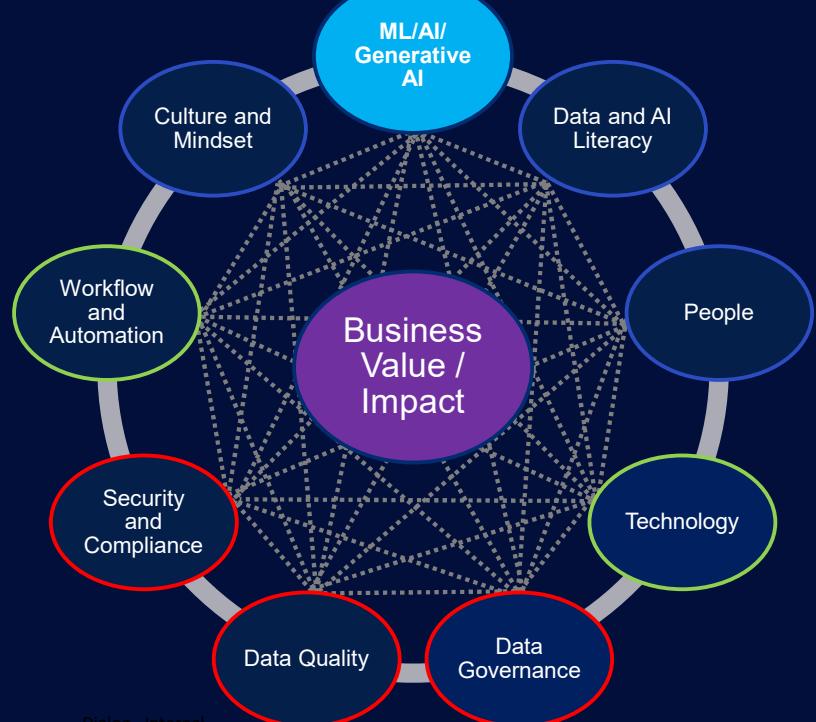


© 2021-2023 Sumudu Tennakoon

Dialog - Internal

41

Putting GenAI To Work



Dialog - Internal

© 2021-2023 Sumudu Tennakoon

42

Challenges related to bringing Generative AI to business

- Data Quality
- Employee Training
- Ethical Considerations
- Security Measures
- Limited Traceability
- Potentially Poor Decision Making
- Investment Priorities
- Lack of Skills
- Biases, Errors, and Limitations
- Dependence on Third-Party Platforms
- Quality of Generated Outputs
- Control Over Outputs
- Computational Requirements
- Bias and Fairness
- Explainability and Interpretability
- Safety and Security

Dialog - Internal

43

Building Applications to Solve (Natural Language) Problems



Dialog - Internal

44

Prompting Best Practices

Set scenario, goal for the chatbot.

Give personas when needed.

Provide Clear, Concise and Unambiguous Instructions as you would provide the same for a Human.

When available, provide supporting data as context or knowledge.

If applicable, Give examples (user prompt and responses): Few shot learning.

Mention restrictions or special instructions to be followed.

- "Do not answer outside the context"
- How to respond when no answer found
- Cite the sources when available.

Specify conditions/limitations

- Output length
- Style
- Creative/Precise
- Audience Control
- Chatbot Level (e.g., 5 year old, creative writer)

Specify Output format

- Structure
- data format like JSON, Tab-delimited, etc.

[Prompt engineering - OpenAI API](#)

© 2021-2023 Sumudu Tennakoon
Dialog - Internal

45

Building AI Applications with Azure OpenAI API

OpenAI Models

- Text Embedding 3:
- GPT 3.5 Turbo (4k, 16k)
- GPT 4: (8k, 32k)
- GPT 4 Turbo (128k)

Embedding

Fine-Tuning*

Prompt Design

Python Libraries

- OpenAI
- Flask
- Streamlit
- LangChain
- Llama-Index

Building Application Components

- Connecting to OpenAI API (Cloud Service)
- LangChain (Chains, Agents and Vector Stores)
- Internal Web API (Application as a Service)
- Web UI (User Interaction)

Build Applications

- Problem Definition -> Prototype -> Evaluation -> Production
- Conversation Chatbot (Your own ChatGPT)
- Chat with own data (Retrieval based Q&A Chatbot)
- Query Assistant (Natural Language to SQL)

© 2021-2023 Sumudu Tennakoon
Dialog - Internal

46

Getting Started with Azure OpenAI

Dialog - Internal

47

Azure OpenAI Models

GPT-4

- A set of models that improve on GPT-3.5 and can understand and generate natural language and code.

GPT-3.5

- A set of models that improve on GPT-3 and can understand and generate natural language and code.

Embeddings

- A set of models that can convert text into numerical vector form to facilitate text similarity.

DALL-E

- A series of models that can generate original images from natural language.

Whisper

- A series of models in preview that can transcribe and translate speech to text.

Text to speech (Preview)

- A series of models in preview that can synthesize text to speech.

<https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/models>

Dialog - Internal

48

Azure OpenAI Service: Models and Prices

Language models			
Models	Context	Input (Per 1,000 tokens)	Output (Per 1,000 tokens)
GPT-3.5-Turbo-0125	16K	\$0.0005	\$0.0015
GPT-3.5-Turbo-Instruct	4K	N/A	N/A
GPT-4-Turbo	128K	\$0.01	\$0.03
GPT-4-Turbo-Vision	128K	\$0.01	\$0.03
GPT-4	8K	\$0.03	\$0.06
GPT-4	32K	\$0.06	\$0.12

Embedding models	
Models	Per 1,000 tokens
Ada	\$0.0001
text-embedding-3-large	\$0.00013
text-embedding-3-small	\$0.00002

<https://azure.microsoft.com/en-us/pricing/details/cognitive-services/openai-service>

49

Creating Azure Resources

Resource Naming Convention:
<https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/ready/azure-best-practices/resource-naming>

The diagram illustrates the hierarchical structure of Azure resources. It shows four levels: Management groups (represented by a group icon), Subscriptions (represented by a key icon), Resource groups (represented by a cube icon), and Resources (represented by various icons like a cloud, code, database, and gear). Dotted lines connect the icons from one level to the next, indicating the relationship between them.

50

Azure OpenAI Service

- Hosted Foundation Models Inference
- Text Embedding
- Fine-tuning (Preview)
- Playground
- Create Assistants
- Deploy LLM Applications

<https://learn.microsoft.com/en-us/azure/ai-services/openai/overview>

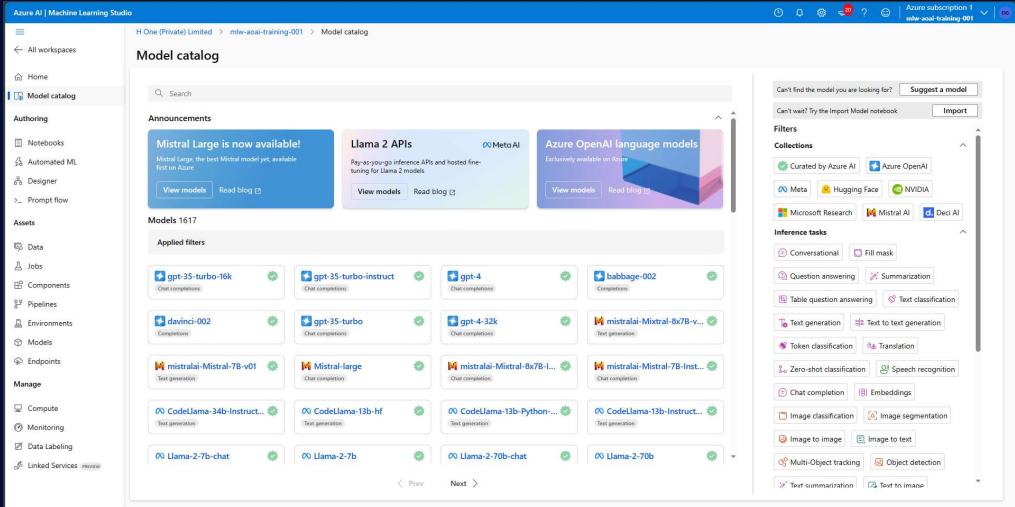
51

Azure OpenAI Studio

The screenshot shows the Azure OpenAI Studio interface in Public Preview. The main header reads "Azure AI | Azure OpenAI Studio PUBLIC PREVIEW". Below it, a banner says "Presenting the new Azure AI Studio (Preview)" and "Build, evaluate, and deploy your AI solutions from end to end." A sidebar on the left lists navigation options: Azure OpenAI, Playground, Chat, Completions, DALL-E, Assistants (Preview), Management, Deployments, Models, Data files, Quotas, and Content filters (Preview). The main content area is titled "Welcome to Azure OpenAI service" and features a "Get started" section with five cards: "Assistants playground (Preview)", "Bring your own data", "Chat playground", "Completions playground", and "DALL-E playground (Preview)". Below this, there's a "Try out common examples" section with four cards: "Customer support agent", "Writing assistant", "Summarize an article", and "Create cover art". At the bottom, a URL is provided: <https://oai.azure.com/portal>.

52

LLM Application Building with Azure ML Studio



The screenshot shows the Azure Machine Learning Studio interface. On the left, there's a sidebar with navigation links like 'All workspaces', 'Home', 'Model catalog', 'Authoring' (Notebooks, Automated ML, Designer, Prompt flow), 'Assets' (Data, Jobs, Components, Pipelines, Environments, Models, Endpoints), and 'Manage' (Compute, Monitoring, Data Labeling, Linked Services). The main area is titled 'Model catalog' and displays a grid of 16 LLM models. Some cards have 'View models' and 'Read blog' buttons. A sidebar on the right lists 'Collections' (Curated by Azure AI, Azure OpenAI, Meta, Hugging Face, NVIDIA, Microsoft Research) and 'Inference tasks' (Conversation, Fill mask, Question answering, Summarization, Table question answering, Text classification, Text generation, Text to text generation, Token classification, Translation, Zero-shot classification, Speech recognition, Chat completion, Embeddings, Image classification, Image segmentation, Image to image, Object detection, Text summarization, Text to image). At the bottom, it says 'Dialog - Internal'.

53

Azure OpenAI Service Samples

<https://github.com/Azure-Samples/openai>



The screenshot shows a GitHub repository page for 'Azure-Samples/openai'. The repository has 1.1k stars and 1.1k forks. It contains 10 branches (main, v1.0, v1.1, v1.2, v1.3, v1.4, v1.5, v1.6, v1.7, v1.8) and 10 commits. The repository description is 'A collection of samples for Azure OpenAI Service, including LLMs, DAVINCI, and CodeLLMs.' The codebase is in Python and .NET. There are sections for 'Code Samples', 'API Reference', and 'Contributing'. At the bottom, it says 'Dialog - Internal'.

54

Thanks!

Do you have any questions?

stennakoon360@gmail.com

www.datasciencefoundations.com

www.linkedin.com/in/sumudutennakoon



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

Dialog - Internal