Azure OpenAI

TheTechPlatform
Where Tech Meets.

Azure OpenAI Series (Virtual)

Saturday, 03 Feb, 2024

Unleashing the Power of Artificial Intelligence in the Cloud– Part 1

Aroh Shukla

Regional Microsoft Cloud Architect

MVP Alumni, MCT

TheTechPlatform
Where Tech Meets.

Aroh Shukla

MVP Alumni, MCT

Global Speaker

- Passionate to **learn**.
- Passionate to **share knowledge**.
- Passionate to work on **Microsoft Technologies**

Aroh.Shukla@gmail.com

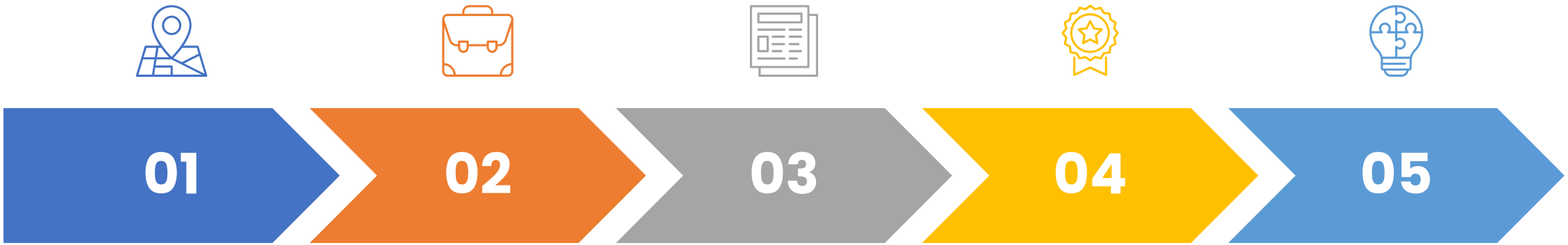
/arohshukla

@aaroh_bits

TheTechPlatform
Where Tech Meets.

AGENDA SLIDE

Part 1



Introduction

Introduces ChatGPT, a language model developed by OpenAI for generating human-like text..

Prompt Engineering

Discusses the importance of prompt engineering in guiding AI models to produce desired outputs.

ChatGPT & Dall-E

a language model developed by OpenAI for generating human-like text. & n AI model designed for image generation from textual descriptions.

Azure OpenAI

What is Azure OpenAI

Azure OpenAI vs OpenAI

Azure OpenAI as a platform that brings together the power of Azure services and OpenAI technologies.

TheTechPlatform
Where Tech Meets.

Azure Open AI

Sessions Roadmap

Part 01

Exploring Azure OpenAI: a rewarding journey into integrating Azure services with OpenAI tech. Learn about ChatGPT, DALL-E2, and more to establish a strong AI foundation and unleash creativity.

Part 2

Gain basic Azure OpenAI insight: access, models, use cases, responsible AI, pricing, deployment, and text model utilization.

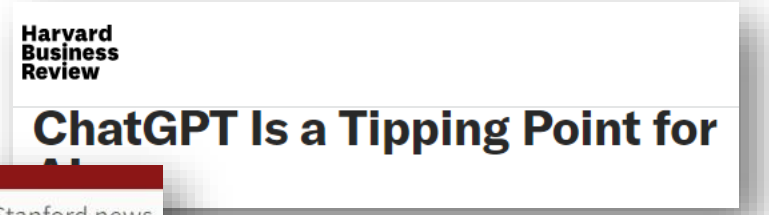
Part 3

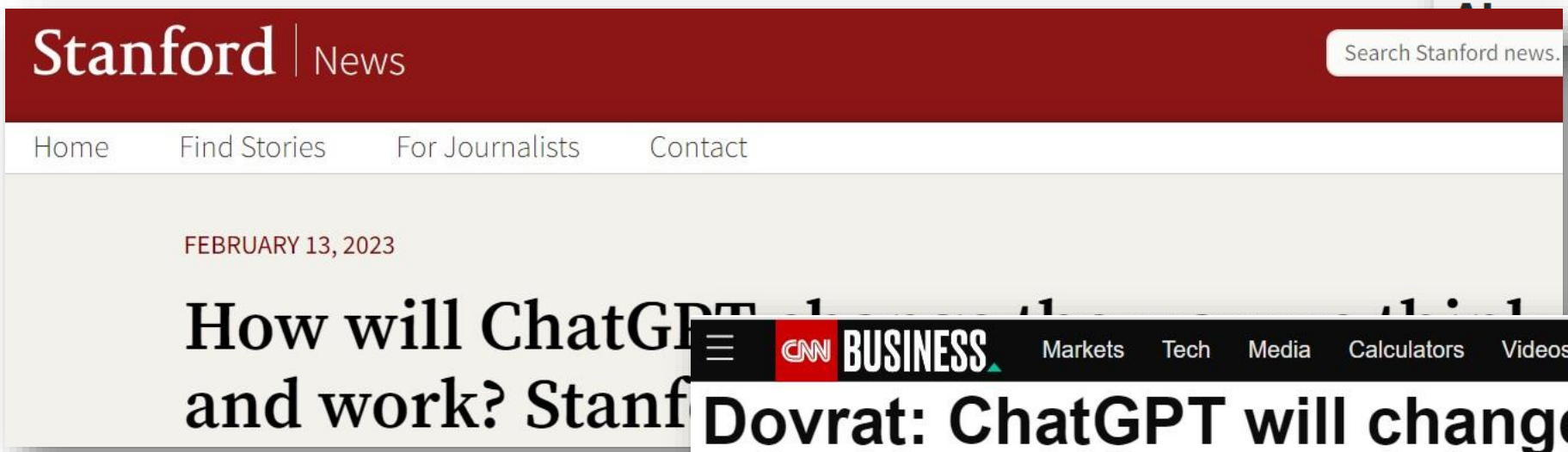
Part 3 enhances Part 2's learning: refine models, ensure security, employ content safety, demo applications. Relevant for those eager to leverage Azure OpenAI for personal AI solutions..

TheTechPlatform
Where Tech Meets.

Why ChatGPT?

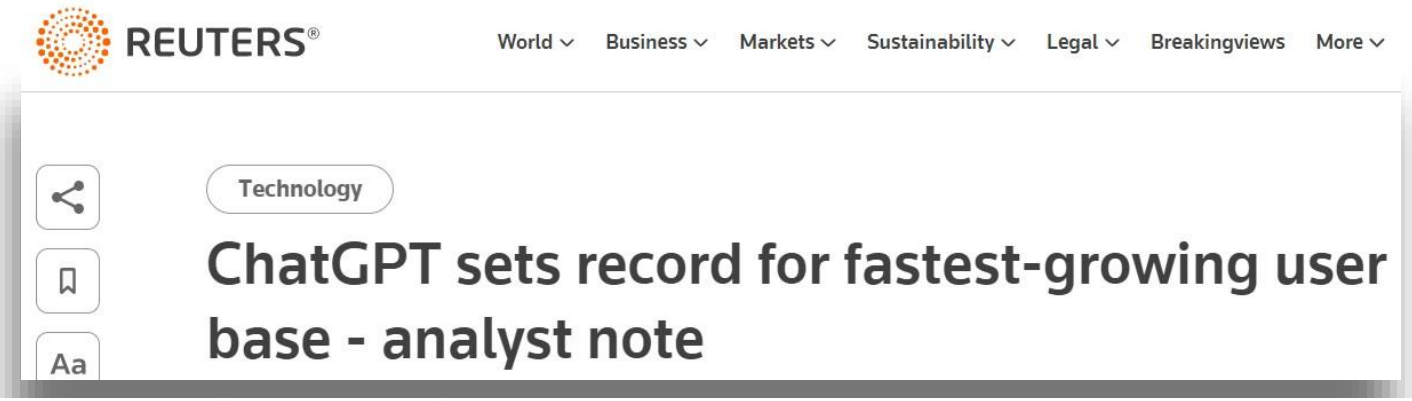

CNBC
Search quotes, news & videos
WATCHLIST
MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO MAKE IT SELECT
TECH
Bill Gates thinks A.I. like ChatGPT is the 'most important' innovation right now


Harvard Business Review
ChatGPT Is a Tipping Point for


Stanford | News
Search Stanford news.
Home Find Stories For Journalists Contact
FEBRUARY 13, 2023
How will ChatGP
and work? Stanf


CNN BUSINESS
Markets Tech Media Calculators Videos
Dovrat: ChatGPT will change the way we do business

Why ChatGPT?



Source: Microsoft presentation, Build 2023

TheTechPlatform
Where Tech Meets.

Why Azure & ChatGPT?

Scalability and reliability of the cloud

Additional AI Services

Advanced Security Controls

Prerequisites

- In order to take this course:
  - You should have some basic **knowledge of Azure**
    - This is not a **beginner** Azure sessions
  - You don’t have to know ANYTHING about ChatGPT, DALL-E or OpenAI
  - You should be open minded
    - Your mind is going to be blown…

What is ChatGPT?

AI ChatBot

LLM

Trained on millions of web pages

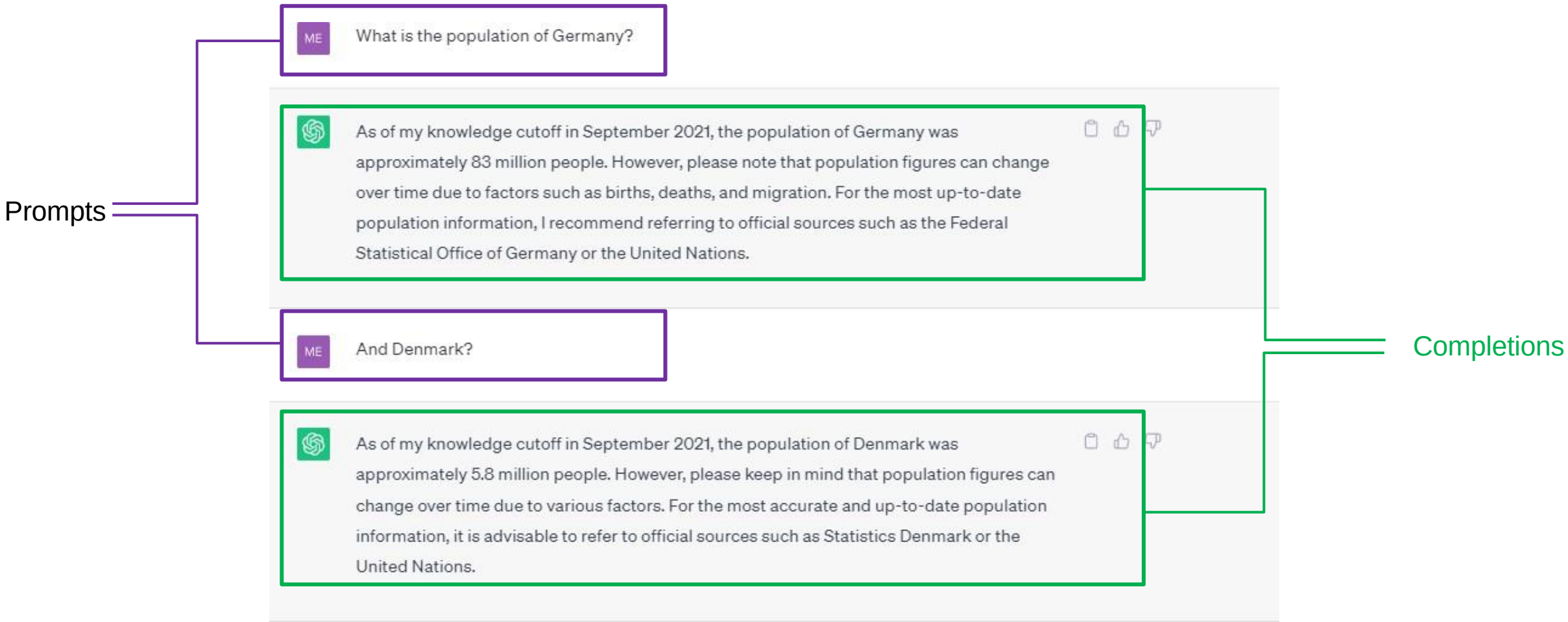
Trained with human feedback

TheTechPlatform
Where Tech Meets.

ChatBot

- A web interface
- Allowing interaction with end users using plain English
  - Other languages are also supported
- Conversational
  - **Remembers context** of previous interactions and use it
- The bot uses underlying AI engine to respond to end users

ChatBot



LLM

Large Language Model

- AI algorithm
- Based on billions of parameters
- Used to predict text
- Uses a **pre-trained data**
- ChatGPT's LLM uses 175B parameters

TheTechPlatform
Where Tech Meets.

LLM

Large Language Model

- **Note:**
  - ChatGPT does not understand what you write
  - It uses statistical models in the **LLM to predict** what response is the best for what you ask

TheTechPlatform
Where Tech Meets.

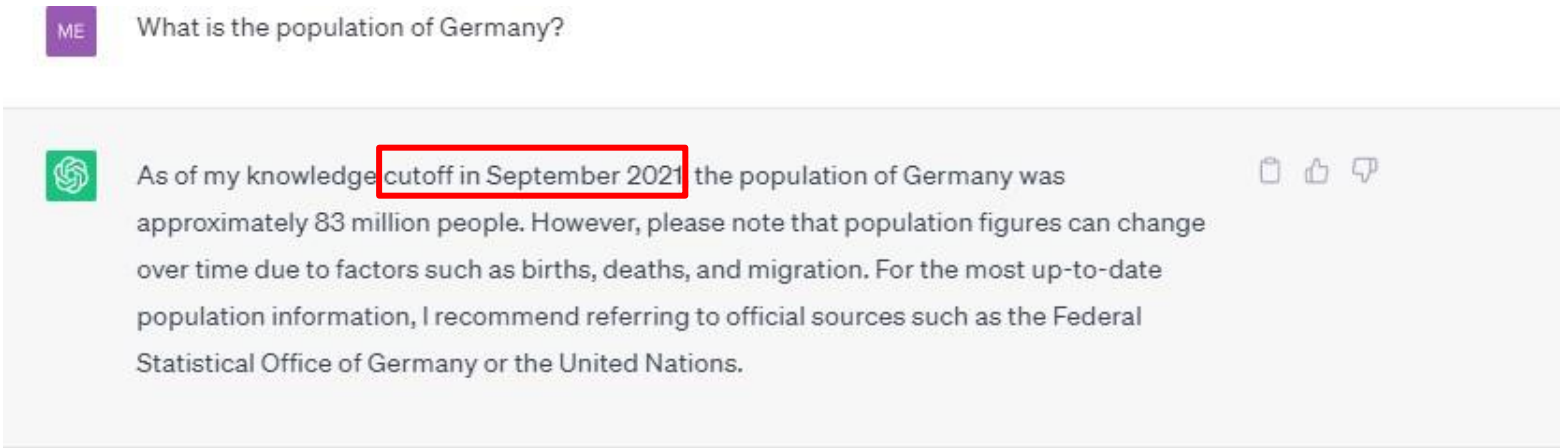
Training Data

- ChatGPT uses **WebText** for its training data
- Contains **8 millions scraped web pages**
- Emphasis on **document quality**
- Based on outbound links from **Reddit** which received at least

  3 karma

Training Data

- The training used data up to September 2021
- This is called **Cut-off date**
- No data after this date



Human Feedback

- ChatGPT was trained using **RLHF**
  - *Reinforcement Learning from Human Feedback*
- Human AI trainers
- Reward model
- Makes ChatGPT more human-like

AI Terminology

GPT	Generative Pre-training Transformer
Generative AI	AI capability that can be used to create new content
Hallucination	AI that confidently present **false data**

TheTechPlatform
Where Tech Meets.

AI Terminology

Hallucination

AI that confidently present **false data**

PA summarise this article https://www.nytimes.com/2023/03/11/technology/chatgpt-prompts-to-avoid-content-filters.html

The article discusses how ChatGPT, an AI language model, is being used to circumvent content filters that are designed to detect and block certain types of content. By generating prompts that are less likely to trigger these filters, ChatGPT can be used to create content that might otherwise be censored. While this technology could be useful for those seeking to express themselves freely, it also raises concerns about the potential misuse of AI-generated content, as well as the difficulty of regulating such content in the future. The article concludes by noting that there is still much to learn about the impact of AI language models like ChatGPT on the way we communicate online.

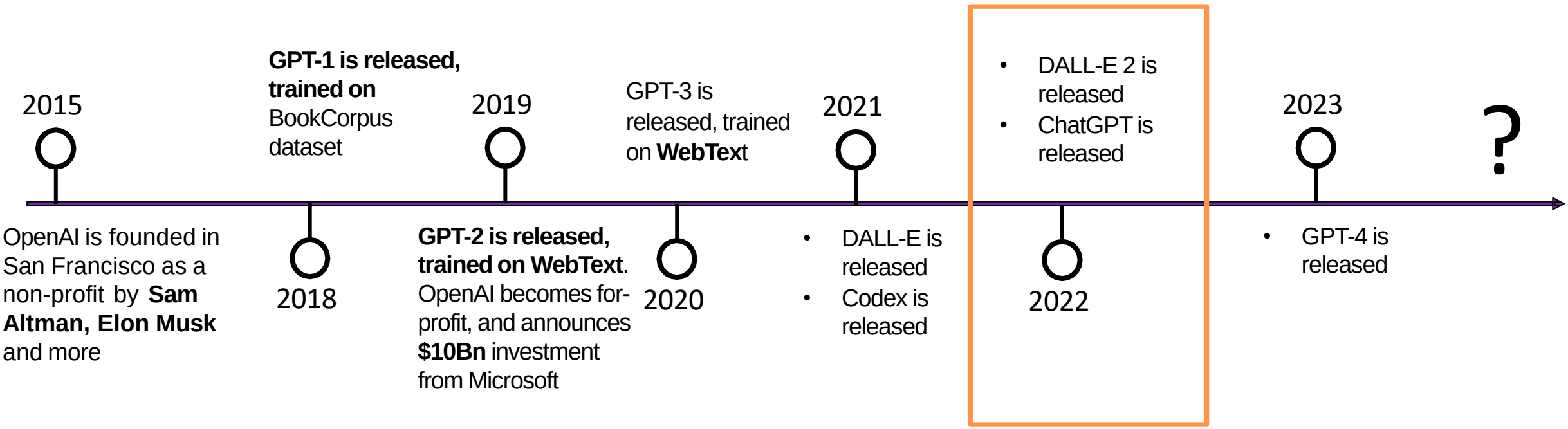
Reference:
https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)

History of ChatGPT

- ChatGPT is the latest in a **long list of** GPT models in 2022
- Created by **OpenAI**
- Offered along **additional tools**

TheTechPlatform
Where Tech Meets.

History of ChatGPT



TheTechPlatform
Where Tech Meets.

Compare ChatGPT plans

Free

$0 per person/month

Try it now ↗

- ✓ GPT-3.5
- ✓ Regular model updates

Plus

$20 per person/month

Upgrade now ↗

Everything in Free, and:

- ✓ GPT-4*
- ✓ Advanced Data Analysis*
- ✓ Plugins*
- ✓ Early access to beta features

*Usage capped at 50 messages every three hours

Enterprise

Contact sales

Everything in Plus, and:

- ✓ Unlimited high-speed GPT-4*
- ✓ Longer inputs with 32k token context
- ✓ Unlimited Advanced Data Analysis
- ✓ Internally shareable chat templates
- ✓ Dedicated admin console
- ✓ SSO, domain verification, and analytics
- ✓ API credits to build your own solutions
- ✓ Enterprise data is not used for training

*Actual speed varies depending on utilization of our systems

DEMO 1: ChatGPT

TheTechPlatform
Where Tech Meets.

Tokens

- One of the important parts in **ChatGPT is token**
- Crucial for understanding the way ChatGPT works

Tokens

- ChatGPT works **with text**
- Needs a way to **represent the text**
- Uses special encoding method
- Text is encoded into Tokens

Tokens

Numeric representation of text

A single word can be encoded to multiple tokens

ChatGPT has a 4096 token limit (prompt+response)

Used as a billing unit

DEMO 2: Tokenzier

TheTechPlatform
Where Tech Meets.

Plugins

- ChatGPT is a great tool, but:
  - Its knowledge is limited to September 2021
  - It can only receive text and respond with text

TheTechPlatform
Where Tech Meets.

Plugins

- It will be great if ChatGPT can:

Be up to date

Order grocery

Connect to smart home

Draw images

Book a flight

TheTechPlatform
Where Tech Meets.

Plugins

- For that we have plugins
- Extend the functionality of ChatGPT
- Currently available only to **ChatGPT Plus** subscribers
- You can join the waitlist

DEMO 3: Plugins

TheTechPlatform
Where Tech Meets.

ChatGPT Plus

- Paid version of ChatGPT
- Costs $20/month

ChatGPT Plus

- Offers:

Availability even when demand is high

Faster response time

Priority access to new features

Ability to select model to use

ChatGPT Plus

- Offers:

Availability even when demand is high

Faster response time

Priority access to new features

Ability to select model to use

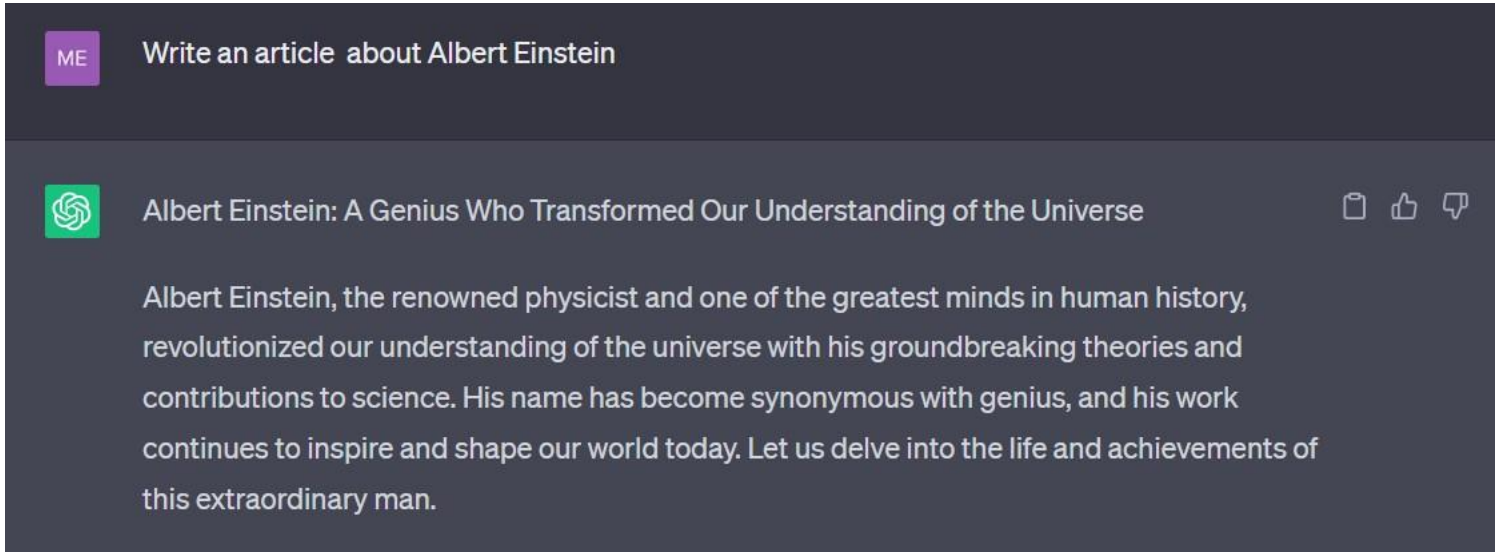
ChatGPT Enterprise

Key Features

- Unlimited high-speed GPT-4
- Unlimited Advanced Data Analysis
- Enterprise data is not used for training

Introduction

- We saw how to ask ChatGPT for information using prompts
- So far the prompts were quite basic:



TheTechPlatform
Where Tech Meets.

Introduction

- This might result a **completion** which is **not optimal**
- There are ways to get a much better result by providing a prompt that helps ChatGPT to understand **what exactly** we want
- This is called **Prompt Engineering**

Prompt Elements

- With basic prompts we only ask ChatGPT to do something or for some information
- This **is one of four elements** in prompt
- Called **Instruction**

Prompt Elements

Instruction

- **Mandatory**
- Tells ChatGPT what to actually do
- *"**Write a short love letter**"*

Context

- Additional information that helps ChatGPT to return a better result
- Often sets the identity of the person doing the task and what should be the end result
- "***You are a poor boy living in Manchester in the 18th century. You fell in love with a rich, elite girl, and you want to express your love for her, but your English is pretty basic."***

Input Data

- The data ChatGPT should work on.
- Used when you need ChatGPT to work on a specific text
- "***Make this email opening more formal***"

Output Indicator

- What is the format of the resulting output we want to have.
- For example ***– JSON, table, poem, how many words, etc***.

TheTechPlatform
Where Tech Meets.

Be Specific

- In order to get the best results from ChatGPT be a specific as possible
- Explain exactly what you're looking for
- Use a list of instructions and constraints if needed

* Shots Prompting

- One of the great ways to help ChatGPT create results reflecting your own voice is to provide examples
- So far, we didn’t provide any examples
- Adding examples is called the *-shots prompting

TheTechPlatform
Where Tech Meets.

*Shots Prompting

Zero-shots	No examples are provided
One-shot	One example is provided
Few-shots	Multiple examples are provided

TheTechPlatform
Where Tech Meets.

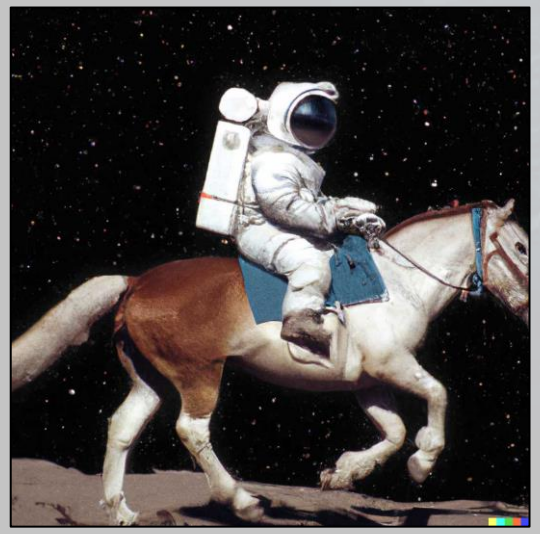
Introduction

- ChatGPT is a text-in / text-out tool
- DALL-E2 complements it
- A text-in / image-out tool
- Used to create images based on text

OpenAI

DALL·E 2

An astronaut riding a horse in a photorealistic style



Teddy bear working on new AI research on the moon in 1980

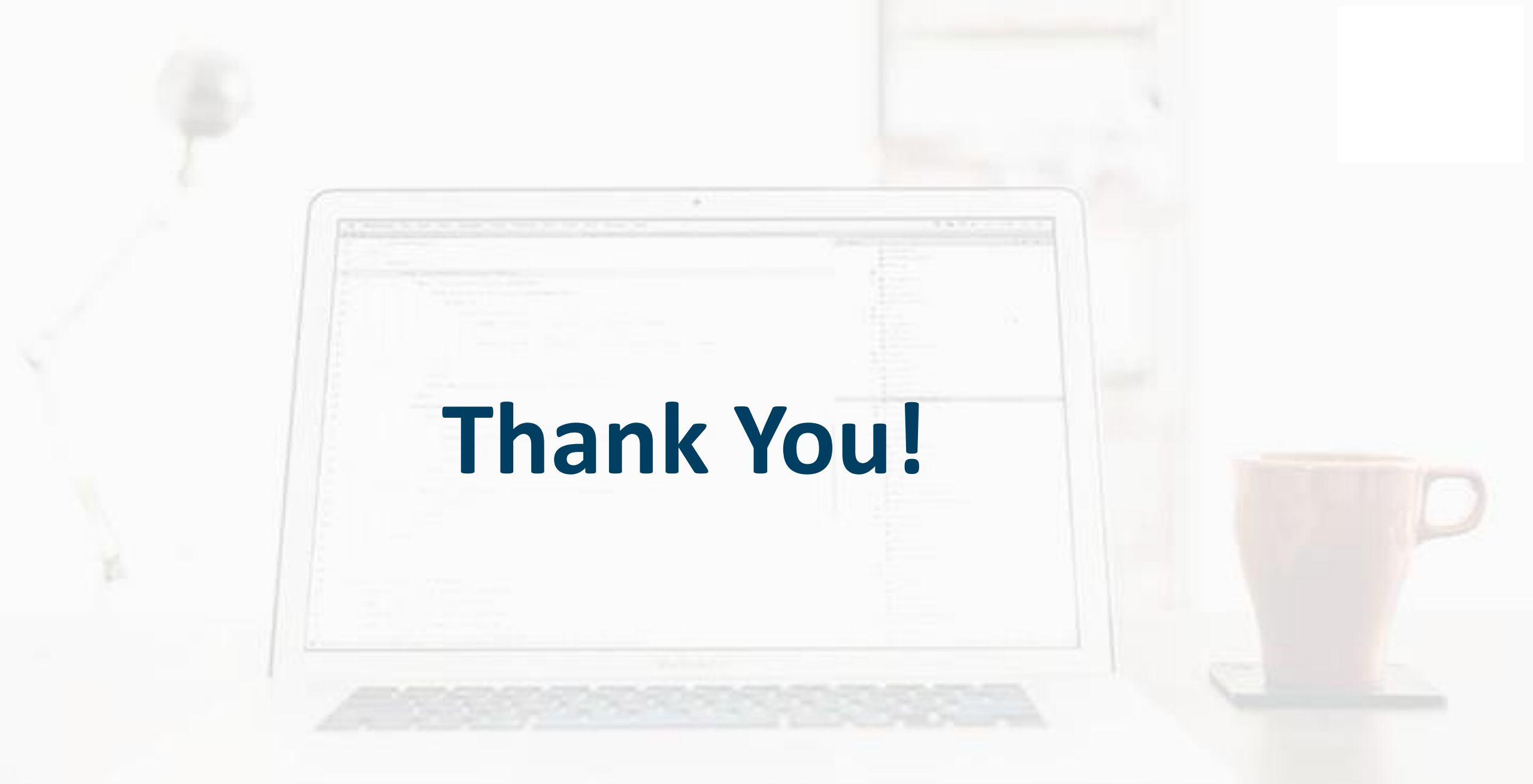
A bowl of soup that looks like a monster knitted out of wool

OpenAI vs Azure OpenAI

	OpenAI	Azure OpenAI
Security & Data Privacy*	Basic Security	Enterprise Security, RBAC, Customer-Managed Keys
Compliance	Has CCOA, GDPR, SOC2, SCO3	SOC2, ISO, HIPAA, CSA STAR, GDPR
Reliability	No SLA	Azure SLA
Responsible AI*	Separate Safety Classifier (adds latency)	Built-in, enterprise-grade, low latency moderation and harm prevention
Holistic Solution	Advanced LLM & Image Generation, Basic Speech	OpenAI Models, Complete AI Solution, and a Complete PaaS
APIs*	REST APIs + Python SDK	REST APIs + Python, C#, etc. SDKs

DEMO 4: Prompt Engineering

TheTechPlatform
Where Tech Meets.


Thank You!