Understanding Big Data & AI

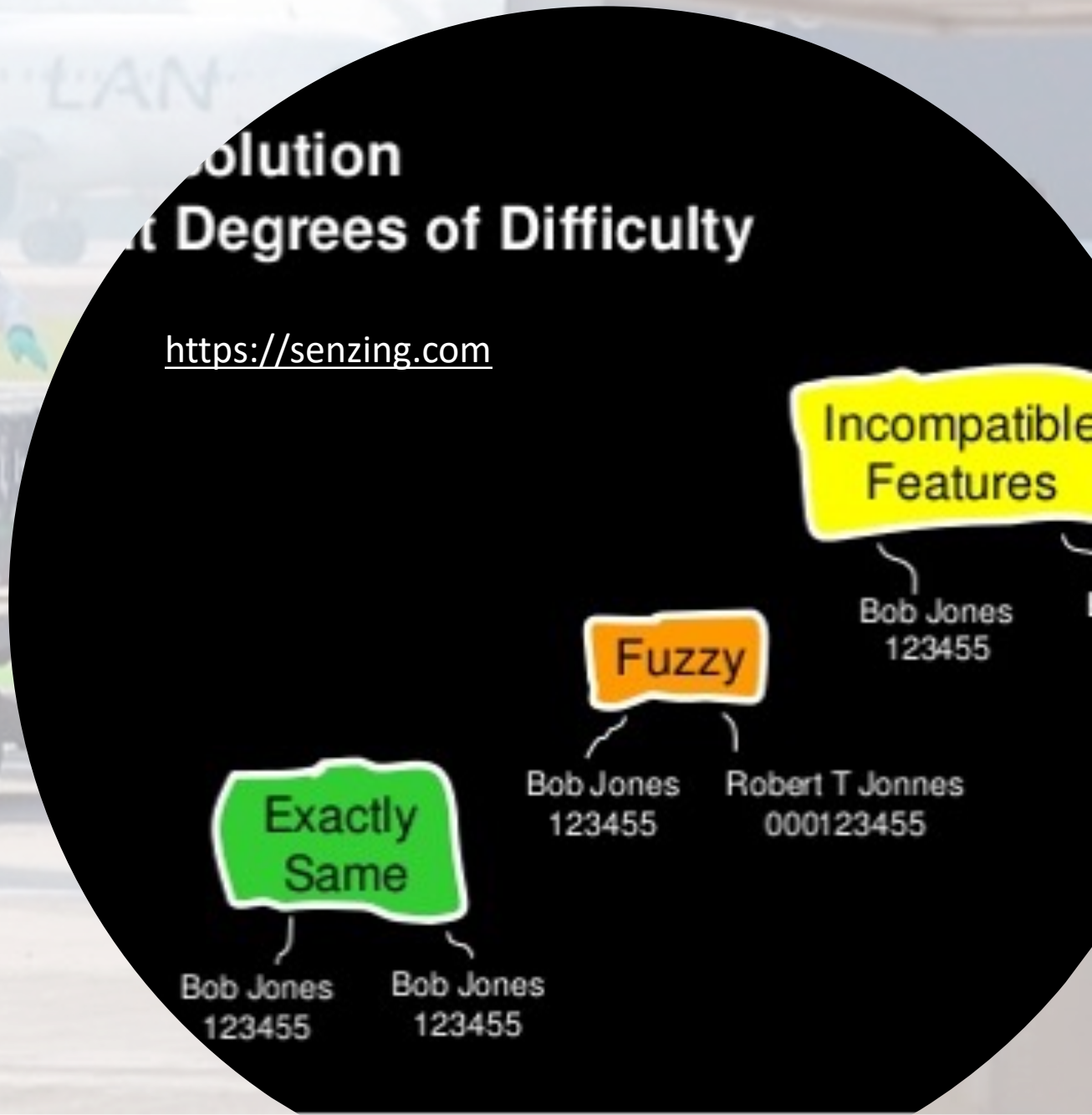
Ivan Portilla 
Portilla@gmail.com

AI Technical Architect
9/26/23

Robo del siglo

https://www.bbc.com/news/world-latin-america-28783027

Entity Resolution



IBM’s Innovation Journey

“Good design is good business.”

Thomas Watson, Jr.
IBM CEO, 1973


IBM
CLR
ON

Be essential.


USA
S-IC-6


IBM
IBM
IBM


MIUM Credit
1234 5678 9012 3456
01/18 01/20


Boarding Pass
Economy Class
PASSENGER
GERHARD PFAU
FLIGHT
EW-2520
FROM
STUTTGART
TO
BARCELO
GATE
A24
TIME
8:12
SEAT
7C
Boarding
Economy Class
PASSENGER
GERHARD PF
TIME
EUROPÄISCHE UNION
BUNDESREPUBLIK
DEUTSCHLAND

Agenda

Big Data & AI

AI Essentials Framework

Demo

Agenda

Big Data & AI

Generative AI (ChatGPT)

Demo

Agenda

Big Data & AI

AI Essentials Framework

Demo

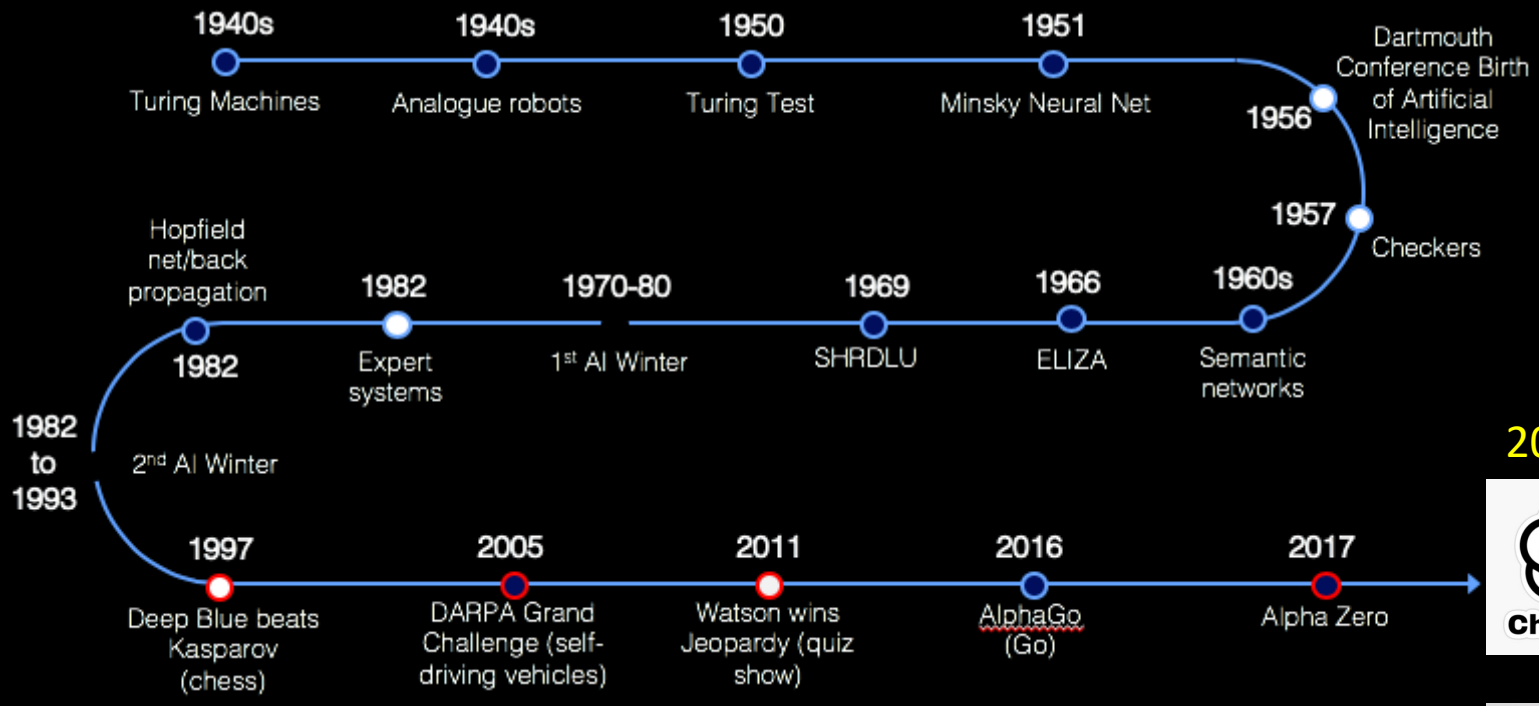
"Every day, we create 2.5 quintillion bytes of data"

Estimates are that less than 0.5% of data is ever analyzed!

Understanding Big Data

Term	Factor	# RAMACs	# IPODs
1 Giga (GB)	10^9	200	
1 Tera (TB)	10^12	200K	200
1 Peta (PB)	10^15	200M	200K
1 Exa (EB)	10^18		200M
1 Zetta (ZB)	10^21		
?	10^24		
?	10^100		

A brief history of AI



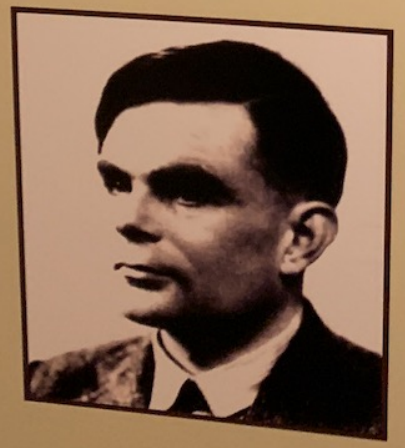
ALAN TURING'S OFFICE

Here at his desk in Hut 8, Turing took the lead on breaking naval Enigma ciphers – something few thought could ever be done. His mathematical skills also enabled him to break other ciphers, including the complex Lorenz cipher where he used a method that became known as Turingery. Together with his fellow Codebreaker Gordon Welchman, he developed the Bombe machine to help speed up the codebreaking process.

IN THEIR WORDS

If anyone was indispensable to Hut 8 it was Turing. The pioneer work always tends to be forgotten when experience and routine later make everything seem easy, and many of us in Hut 8 felt that the magnitude of Turing's contribution was never fully realised by the outside world.

Hugh Alexander, Codebreaker, Hut 8



What is AI?

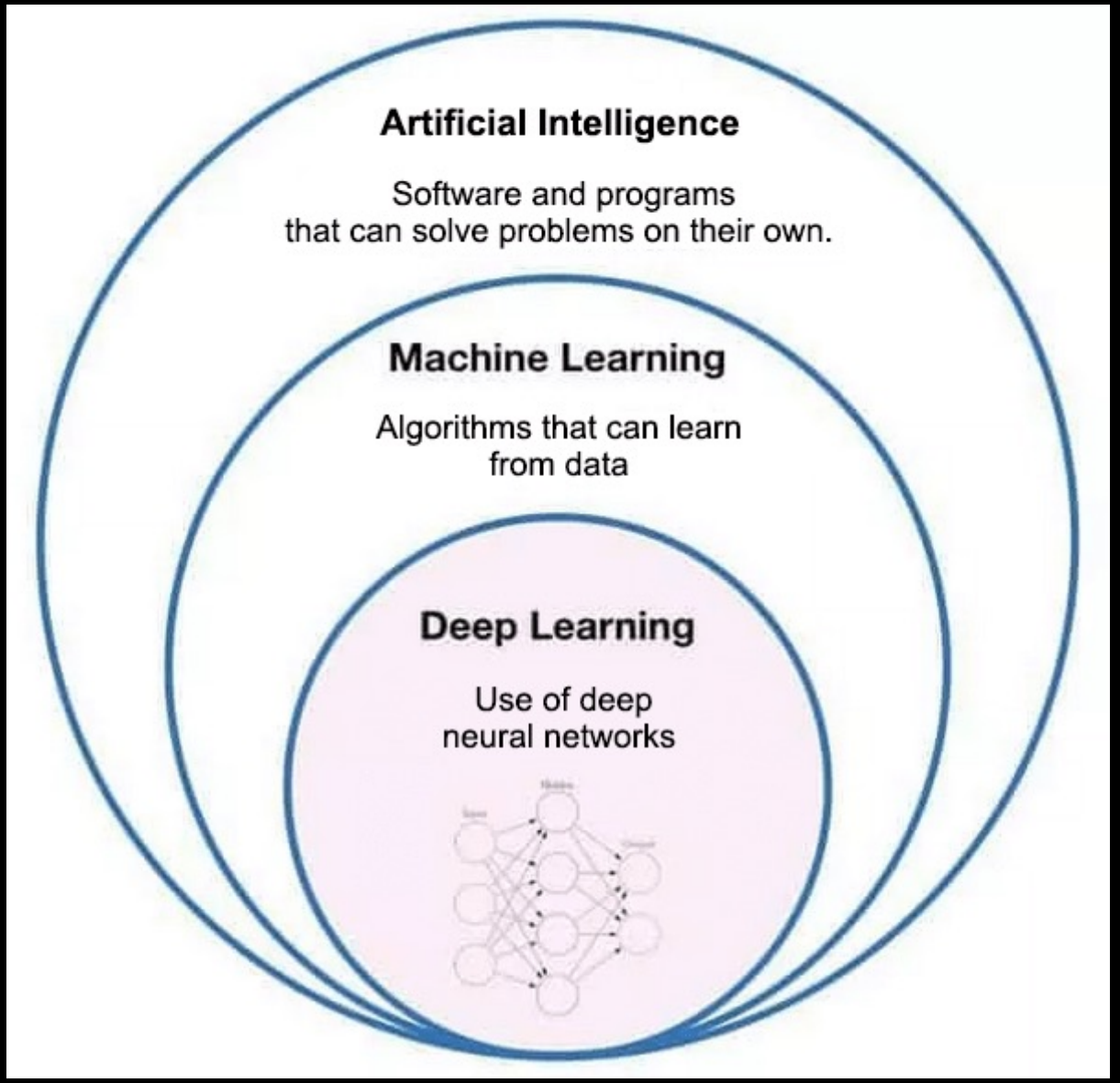
AI Perception

What do people feel about AI?

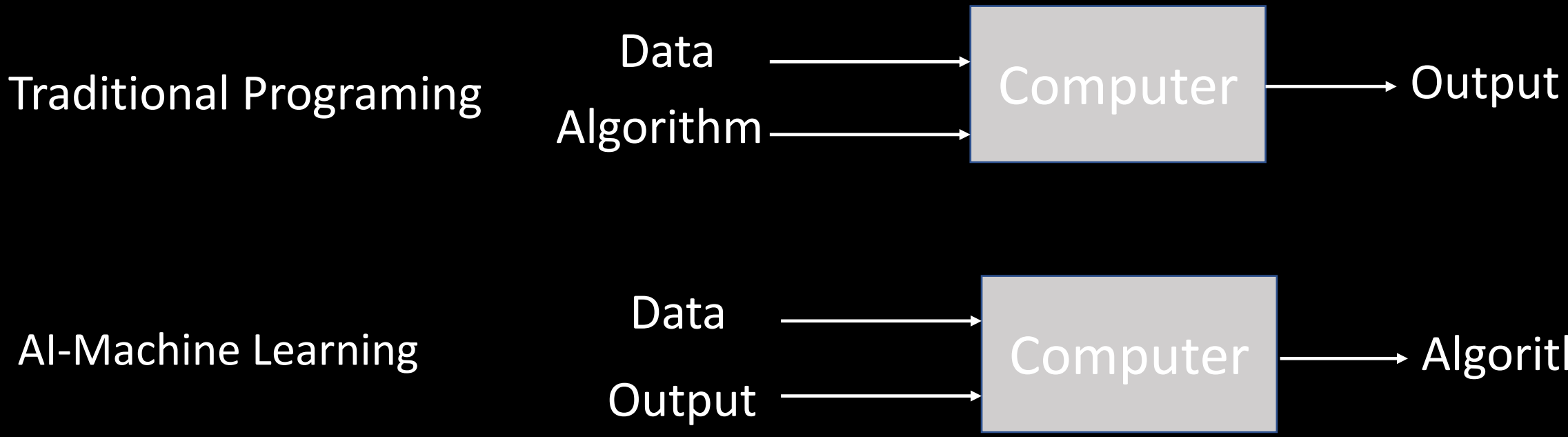
Fearful
Skeptical
Fascinated
Understood

Exposed
Apprehensive
Curious
Excited
Well-served

What we know


Artificial Intelligence
Software and programs
that can solve problems on their own.
Machine Learning
Algorithms that can learn
from data
Deep Learning
Use of deep
neural networks

Why Machine Learning?



ML provides systems the ability to automatically learn from experience
(without being explicitly programmed)

Intelligence demonstrated by CU Students

Exercise One


PB&J Recipe
wikiHow to Make a Peanut Butter and Jelly Sandwich

Exercise 1

Program how to make a Peanut Butter & Jelly sandwich

1. Get ingredients (PB, Jelly, Bread)
2. Get cooking utensils (plate, knives)
3. Get two slices of bread
4. Open PB, Jelly jars
5. With a knife spread 2 Tablespoons of peanut butter on one piece of bread
6. With a different knife, spread 2 Tablespoons of jelly on the other slice of bread
7. Put the slices together
8. Toddler adaptation: cut off crusts before serving.
9. Place on plate & serve

Exercise Two


Passing truck in highway

Passing truck in highway

Don't Follow Closely Behind Prior to Passing

- As you prepare to pass, keep a healthy distance between your car and the truck in front of you. At minimum, 30 feet should separate the two vehicles. Signal clearly to indicate your next move.

Pass in the Left Lane

- The left lane is referred to as the 'passing lane' for a reason; it's the safest place to pass any vehicle, but particularly trucks and buses. It's frustrating to be stuck behind a slow truck in the left lane, but resist the urge to pass on the right side.

Don't Linger in the Truck's Blind Spot

- Trucks have huge blind spots on both sides. Pass promptly while still abiding by the speed limit. Don't hang out in the blind spot area, which, in the left (passing) lane, constitutes the truck's entire front half.

https://www.driverseducationusa.com/resources/passing/

Recognition

- With GPS navigation, HiRes cameras, Laser-radar cameras: Location, speed of trucks, road conditions, emergency lane
- With Image Recognition: Turn signal (!lamp), paper or debris on the road

Insights

- Data is fed into multiple CPUs + collected data + rules
- Manufacturer knows who has the right of the way, when trucks can pass + MM miles driven have feedback on different driving style

Action (Cognitive Decision)

- Assist (alerts driver)
- Act (Pass or Not pass, Measure decision effects, feedback loop, and share w other cars)

Exercise Three

How is AI being used today?

AI is Everywhere

AUTOMATION	CONVERSATION	OPTIMIZATION	RECOMMENDING AND INSIGHTS	PERSONALIZING	VISUAL RECOGNITION
					
Automating actions, processes, and alerts	Understanding text and answering questions in natural language	Analyzing data for patterns and outliers and deriving insights	Understanding needs and recommending solutions	Personalizing experiences	Recognizing objects in images and videos

What Machine Learning Can Do

A simple way to think about supervised learning.

INPUT A	RESPONSE B	APPLICATION
Picture	Are there human faces? (0 or 1)	Photo tagging
Loan application	Will they repay the loan? (0 or 1)	Loan approvals
Ad plus user information	Will user click on ad? (0 or 1)	Targeted online ads
Audio clip	Transcript of audio clip	Speech recognition
English sentence	French sentence	Language translation
Sensors from hard disk, plane engine, etc.	Is it about to fail?	Preventive maintenance
Car camera and other sensors	Position of other cars	Self-driving cars

SOURCE ANDREW NG

https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now

Future of AI Example (Semantic Analysis)

The Wining Shot

The soccer game was nearly over. The two teams were tied, one to one.

Alice kicked the ball. Oops! She kicked it the wrong way.

”What a bad kick,” Alice thought.

But wait!. Her teammate John jumped to stop the ball. The ball went off John into the goal!.

Alice’s team won the game. Hooray! It was not a bad kick, after all.

Who won the game?

Why?

Future of AI Example

The Wining Shot

The soccer game was nearly over. The two teams were tied, one to one.

Alice kicked the ball. Oops! She kicked it the wrong way.

”What a bad kick,” Alice thought.

But wait!. Her teammate John jumped to stop the ball. The ball went off John into the goal!.

Alice’s team won the game. Hooray! It was not a bad kick, after all.

Who won the game?

Answer: Alice’s team
Correct

Explanation: Alice’s team won the game because Alice’s team had more points than the other team
✓Correct

https://www.oreilly.com/radar/machines-as-thought-partners/

ChatGPT

JE

read this: The Wining Shot

The soccer game was nearly over. The two teams were tied, one to one.

Alice kicked the ball. Oops! She kicked it the wrong way.

"What a bad kick," Alice thought.

But wait!. Her teammate John jumped to stop the ball. The ball went off John into the goal!.

Alice's team won the game. Hooray! It was not a bad kick, after all.

ChatGPT

JE

who won the game

Alice's team won the game, even though Alice initially thought she had made a bad kick. John's quick thinking and intervention helped the team to score the winning goal, and they were able to secure victory in the game.

JE

why?

Alice's team won the game because John was able to use his skills and reflexes to intercept the ball, even though Alice's kick was not aimed in the right direction. John's intervention caused the ball to go into the goal, which resulted in a score for Alice's team. Since this happened near the end of the game and there was no time left for the other team to catch up, Alice's team won the game with a score of 2-1.

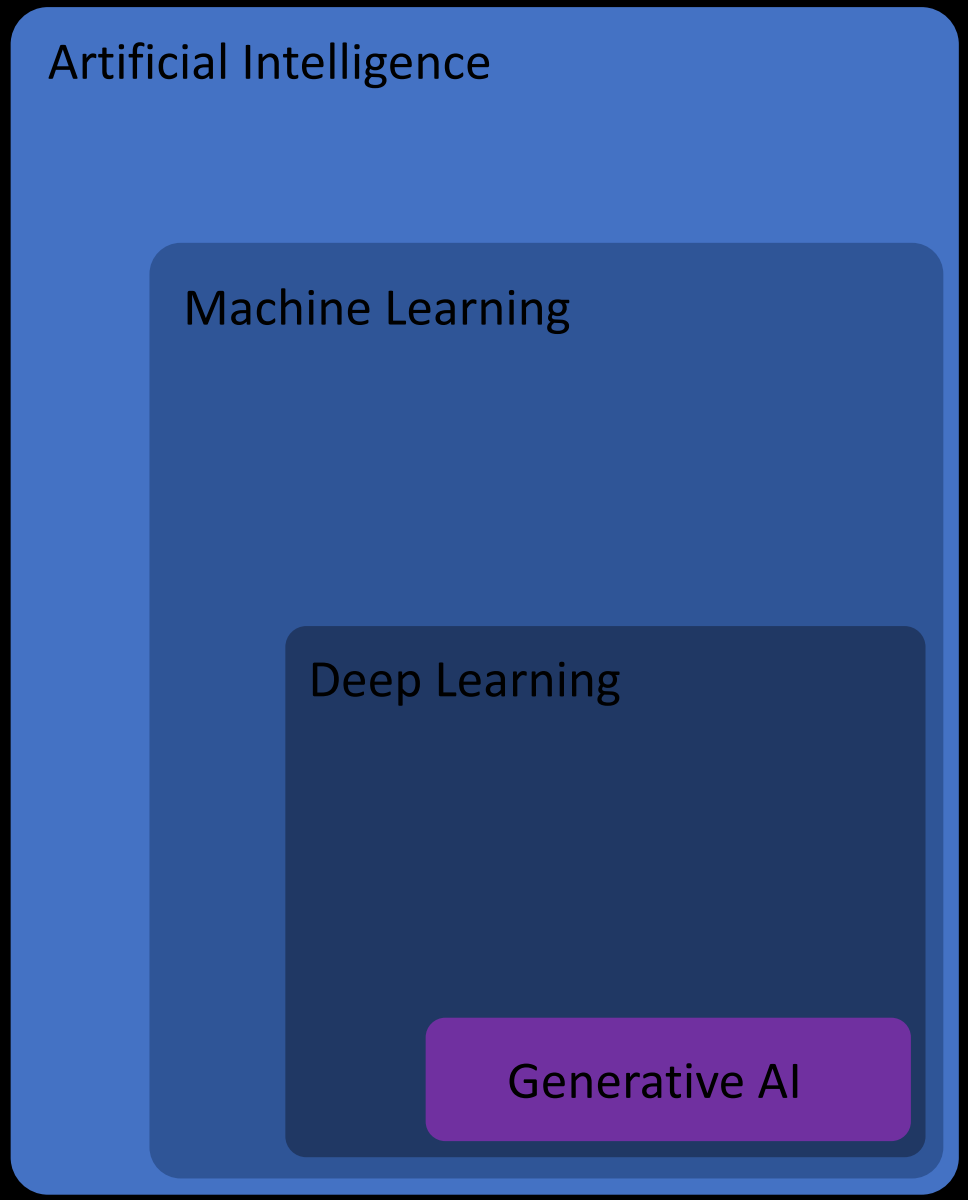
Agenda

Big Data & AI

Generative AI (ChatGPT)

Demo

Brief history of AI



1956

Artificial Intelligence

Intelligence demonstrated by machines

1997

Machine Learning

Learn from data

2017

Deep Learning

Model after the human brain (Neural Networks)

2021

Generative AI

Create new written, visual, and auditory content

MSFT partnership with OpenAI

Ensure that artificial general intelligence (AGI) benefits humanity.

Empower every person and organization on the planet to achieve more

GPT-3

Generate and Understand Text

Codex

Generate and Understand Code

DALL·E

Generate images from text prompts

OpenAI Microsoft

Generative AI

GPT-3

Prompt:

```
Write a tagline for an ice
cream shop.
```

Response:

```
We serve up smiles with every
scoop!
```

Codex

Prompt:

```
Table customers, columns =
[CustomerId, FirstName,
LastName, Company, Address,
City, State, Country,
PostalCode]

Create a SQL query for all
customers in Texas named Jane
query =
```

Response:

```
SELECT *
FROM customers
WHERE State = 'TX' AND
FirstName = 'Jane'
```

DALL·E

Prompt: A white Siamese cat

Response:

Foundation Models

Data

- Text
- Images
- Speech
- Structured Data
- 3d Signals

Training

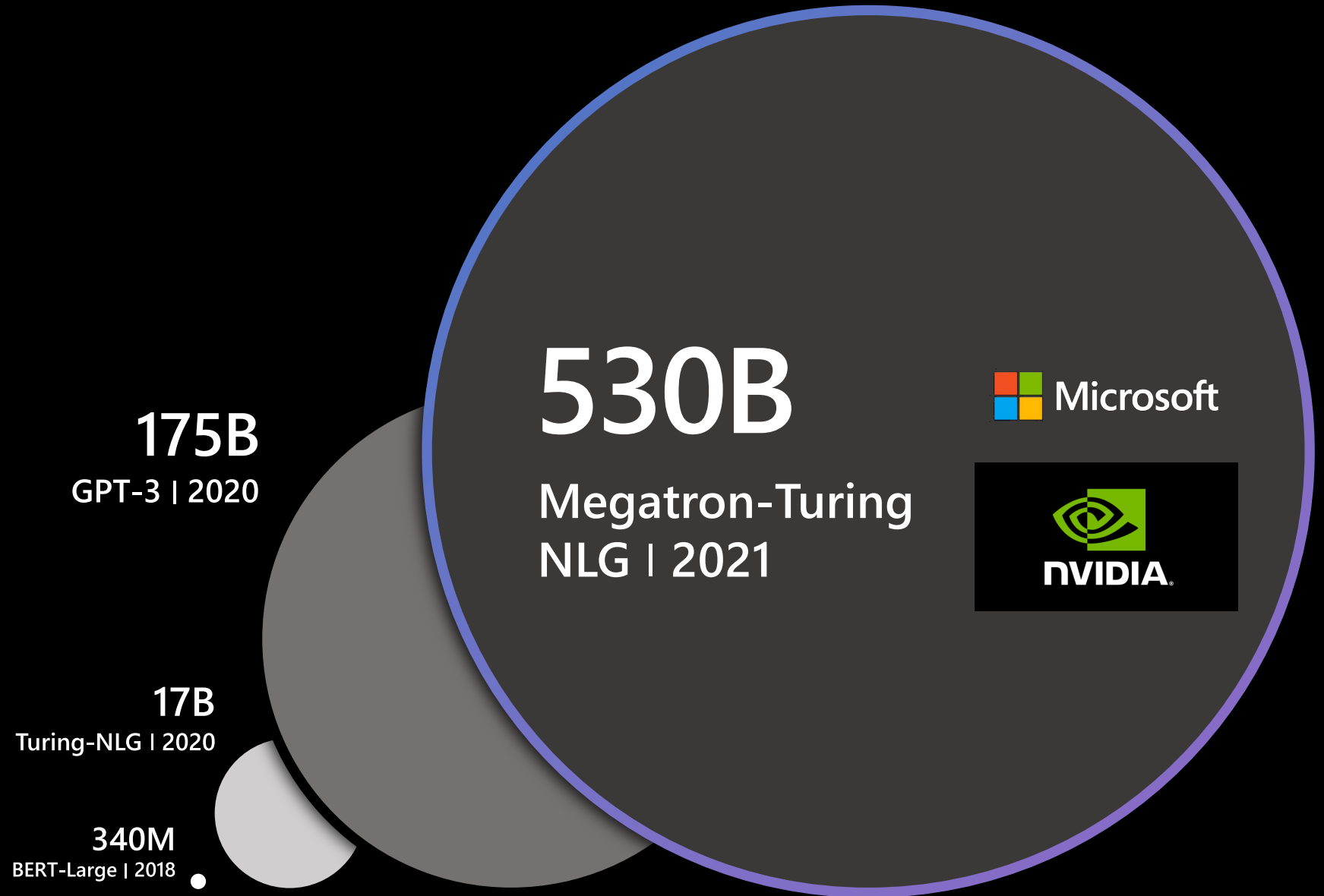
Foundation Model

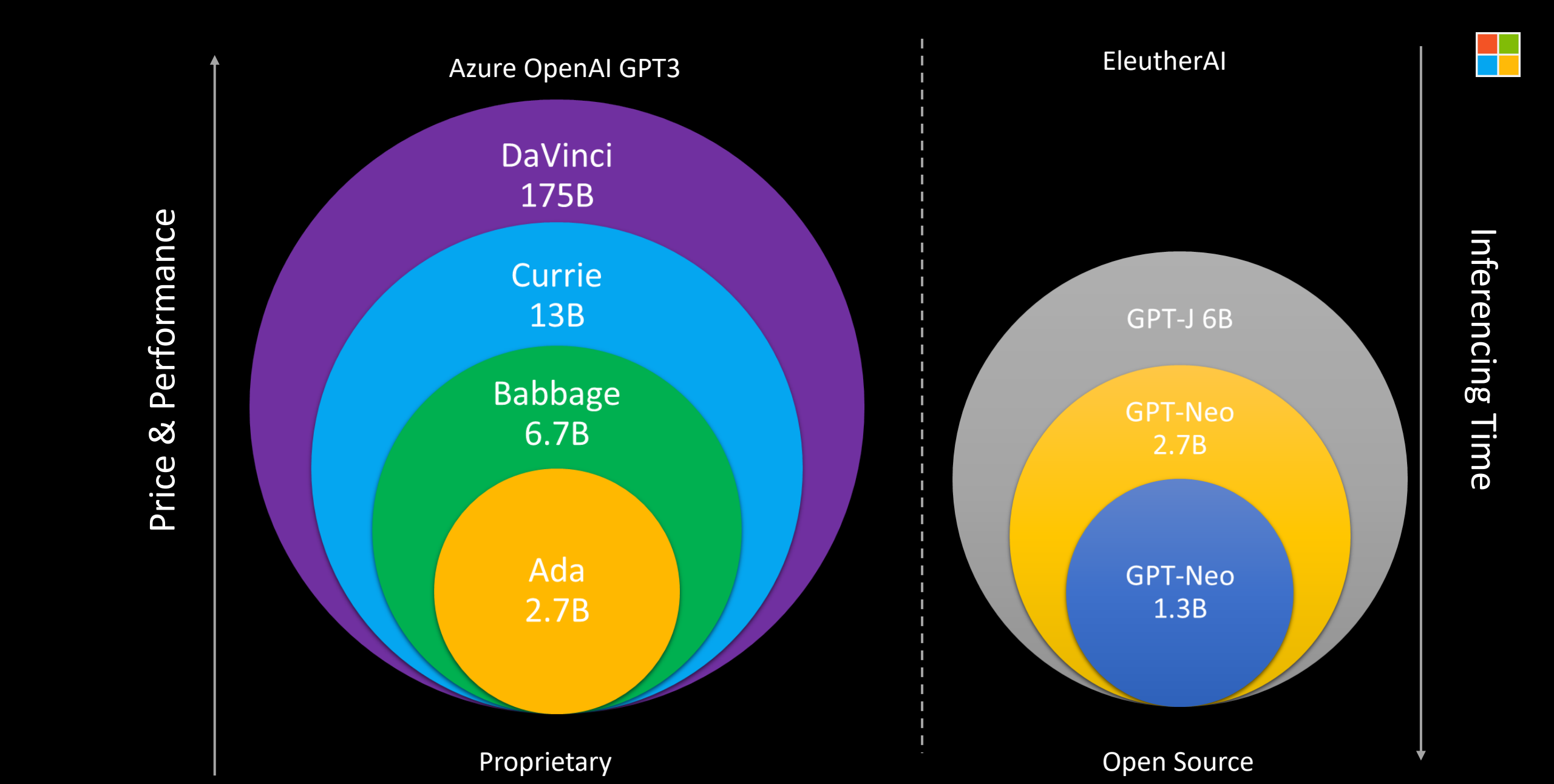
Transformer Model

Adaptation

Tasks

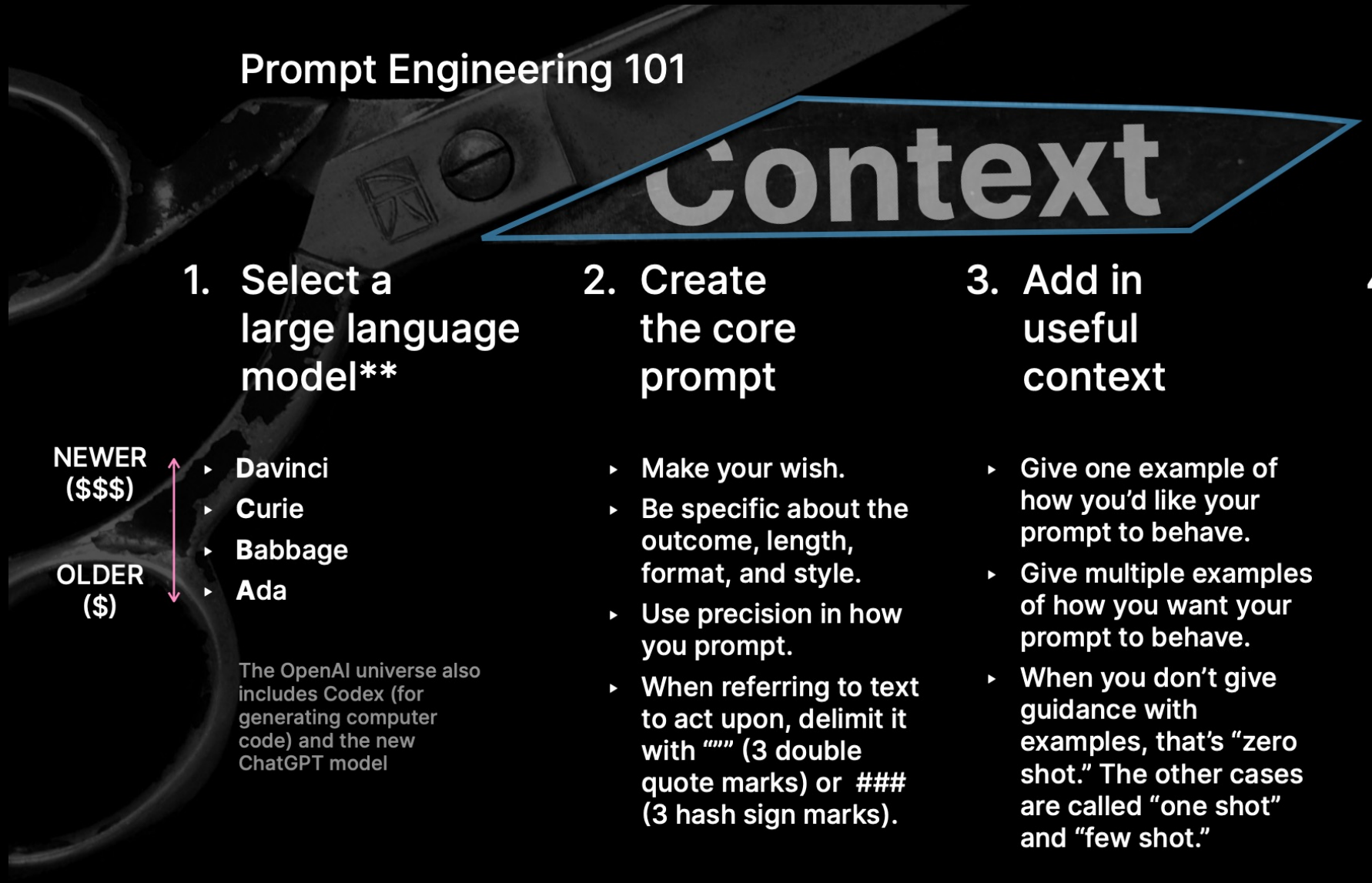
- Question and Answering
- Sentiment Analysis
- Information Extraction
- Image Captioning
- Object Recognition
- Instruction Follow


530B
Megatron-Turing
NLG | 2021
Microsoft
NVIDIA
175B
GPT-3 | 2020
17B
Turing-NLG | 2020
340M
BERT-Large | 2018



Demos

Prompt Engineering 101



1. Select a large language model**

NEWER ($$$)

OLDER ($)

- **D**avinci
- **C**urie
- **B**abbage
- **A**da

The OpenAI universe also includes Codex (for generating computer code) and the new ChatGPT model

2. Create the core prompt

- Make your wish.
- Be specific about the outcome, length, format, and style.
- Use precision in how you prompt.
- When referring to text to act upon, delimit it with “”” (3 double quote marks) or ### (3 hash sign marks).

3. Add in useful context

- Give one example of how you’d like your prompt to behave.
- Give multiple examples of how you want your prompt to behave.
- When you don’t give guidance with examples, that’s “zero shot.” The other cases are called “one shot” and “few shot.”

4. Tune the model’s settings

- Set the temperature high to make it more random; lower the temperature to make it more deterministic.
- Alternatively, adjust the “p” parameter to broaden or narrow the range of words that will be generated in response.
- Give it more tokens when you need to work with longer prompt or longer responses.

Text Classification

Named Entity Recognition (Zero-Shot)

Extract the name and mailing address from this email:

Dear Kaylee,
It was great to talk to you at the seminar. I thought Peter’s talk was quite good.
Thank you for the book. Here's my address 995 Regent Dr, Boulder, CO 80309
Best,
Marilú

Name: Marilú
Mailing Address: 995 Regent Dr, Boulder, CO 80309

Text Classification

Custom Named Entities (Few-Shot)

Extract job titles from the following sentences.

Sentence: Jane Doe has been working for IBM for 20 years as a Linux Engineer.
Job title: Linux Engineer
###
Sentence: Jane Doe has been working for IBM for 20 years and she loved it.
Job title: none
###
Sentence: Demis Hassabis| Director | Google
Job title: Director
###
Sentence: Juan is the CTO of Platform.sh, he was previously the CTO of Commerce Guys, a leading ecommerce provider.
Job title: CTO

Text Classification

Multi-Label Text Classification (Few-Shot)

The following is a list of movies and the categories they fall into:

Topgun: As students at the United States Navy's elite fighter weapons school compete to be best in the class, one daring young pilot learns a few things from a civilian instructor that are not taught in the classroom.
Category: Drama, Action

Hustle: A basketball scout discovers a phenomenal street ball player while in Spain and sees the prospect as his opportunity to get back into the NBA.
Category: Comedy, Drama, Sport

The Northman: From visionary director Robert Eggers comes The Northman, an action-filled epic that follows a young Viking prince on his quest to avenge his father's murder.
Category: Action, Adventure, Drama

Reason Over Unstructured Text

Question and Answering

On Monday, Microsoft launched an OpenAI service as part of its Azure cloud platform, offering businesses and start-ups the ability to incorporate models like ChatGPT into their own systems. The company has already been building AI tools into many of its consumer products, such as a DALL-E 2 feature in its Bing search engine that can create images based on a text prompt, and the Information reported recently that it’s working to bring more of them to Microsoft Office as well.

What will Microsoft incorporate in its own systems?

Microsoft plans to incorporate AI tools into its own systems, such as a DALL-E 2 feature in its Bing search engine that can create images based on a text prompt, and the Information reported recently that it’s working to bring more of them to Microsoft Office as well

Summarization

Extractive vs Abstractive

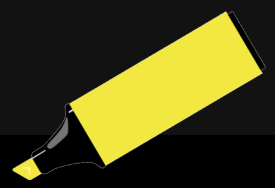
Extractive

Select parts of the original text to form a summary

- Easier
- Restrictive (no paraphrasing)

Peter and Elizabeth took a taxi to attend the night party in the city. While in the party, Elizabeth collapsed and was rushed to the hospital.

Peter and Elizabeth attend party city. Elizabeth rushed hospital



Abstractive (Semantic)

Generates new text using natural language generation techniques

- More difficult
- More flexible (more human)

> Peter and Elizabeth took a taxi to attend the night party in the city. While in the party, Elizabeth collapsed and was rushed to the hospital.

Elizabeth was hospitalized after attending a party with Peter.

Abstractive Summarization

Contact Center Summarization

Summarize this for a call center agent:

Agent: Thank you for calling ADM. Who am I am speaking to?
Customer: Hello, my name is Jeremiah Contreras. I own a small business and have some questions regarding payroll processing.
Agent: Good morning, Jeremiah, before we get started may I ask you a few questions so that I better answer your questions today?
Customer: Thank you that is quite helpful. Are there specific regulations that I need to follow?
Agent: Certain aspects of payroll processing are regulated by the Internal Revenue Service (IRS) and the Department of Labor (DOL)

The agent is speaking to a customer named Jeremiah Contreras. Jeremiah is asking for tips on manual payroll processing. The agent also informs Jeremiah that certain aspects of payroll processing are regulated by the Internal Revenue Service (IRS) and the Department of Labor (DOL).

Prompt Insert

Only DaVinci-002+

Before

These are agenda topics for a customer presentation.

1. Introduction to GPT3
[insert]
10. Conclusion

After

These are agenda topics for a customer presentation.

1. Introduction to GPT3
2. Overview of GPT3 Features
3. Benefits of GPT3
4. GPT3 Use Cases
5. GPT3 Pricing
6. GPT3 Security
7. GPT3 Support
8. GPT3 Integration
9. Q&A Session
10. Conclusion

Media Example

News Analyses & Article Creations

News Broadcast (Global Warming)

Azure Speech Transcription

Azure OpenAI Text Completion

Search
SEO
Virtual Agent
Analytics
Reporting
Knowledge Mining

Content Analysis

Content Creation

Ideation
Productivity
Personalization
Accessibility
A/B Testing

Topic Classification

Global warming, Deforestation, Carbon footprint

Entity Extraction

Organizations: IPCC, UNFCCC, Green Peace
Geography: Canada, USA

Key Word Extraction

Human activities, fossil fuels, earth atmosphere

Question and Answer

What is the Intergovernmental Panel on Climate Change (IPCC)?
The IPCC is an international organization that studies climate change and the effectiveness …

Video summarization

The article discusses about global warming and its effects on the Earth's atmosphere, wildlife, and human communities. It states that the primary cause of global warming is ….

News article generation (or blogs, social media)

Global warming is the gradual increase in the overall temperature of the Earth's atmosphere, primarily caused …

Script Generation

Act 1: The show opens with a shot of a beautiful coastal town
Act 2: As the town struggles to cope with the crisis …

Personalized Content generation (or Advertising)

Simon, as someone passionate about global warming, you are aware of the urgent threat it poses to our planet …

Sports & Entertainment Example

Natural language to SQL to surface stats data (Fan Engagement)

"How many points did Lebron James score in 2021?"

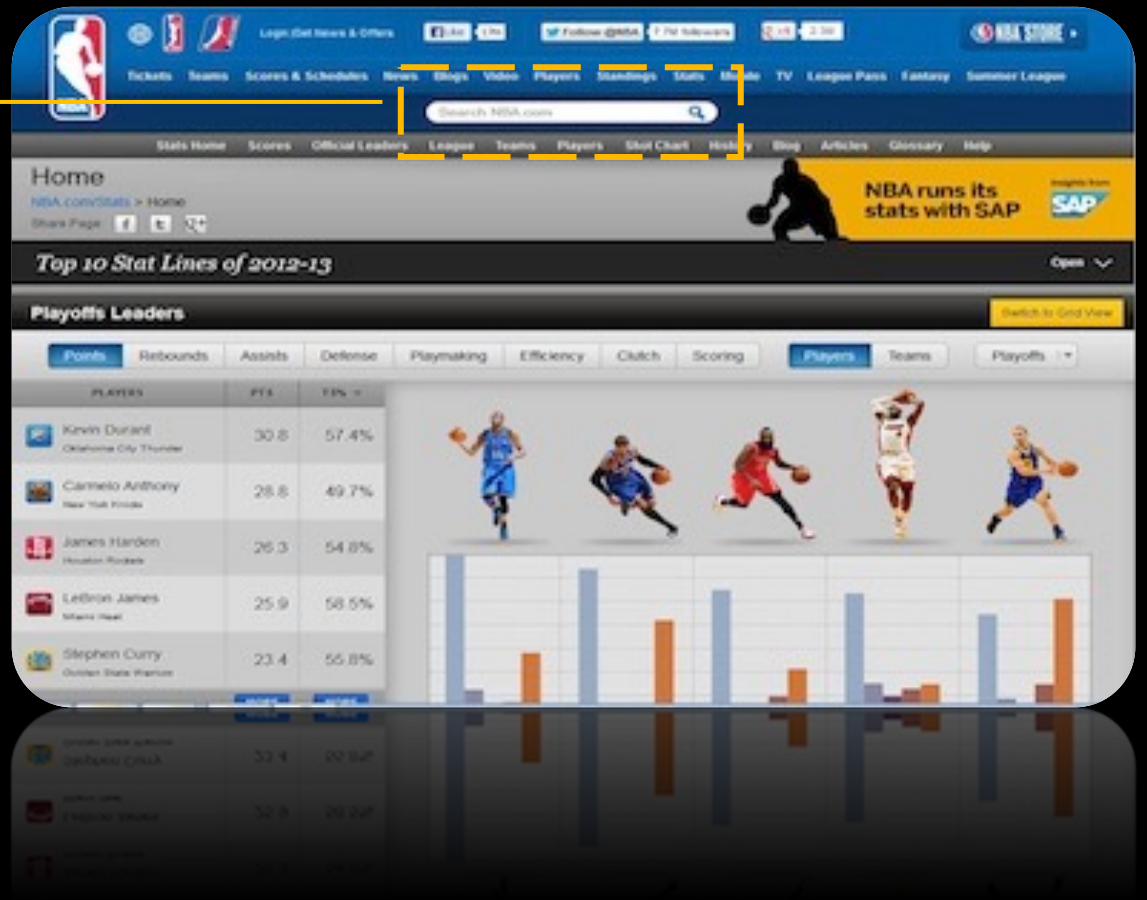
SELECT SUM(points) FROM basketball_stats WHERE player = 'LeBron James' AND game_date BETWEEN '2021-01-01' AND '2021-12-31';

Azure OpenAI Codex

orts Stats
atabase

basketball_stats (player, team, points, rebounds, assists, steals, blocks, turnovers, games_played)
("LeBron James", "Los Angeles Lakers", 25, 7, 10, 1, 0, 3, 82), ("Kevin Durant", "Brooklyn Nets", 30, 7, 4, 1, 2, 3, 72)

basketball_games (home_team, away_team, home_score, away_score, game_date)
("Los Angeles Lakers", "Brooklyn Nets", 110, 105, '2021-12-01'),
("Milwaukee Bucks", "Golden State Warriors", 120, 115, '2021-12-02')



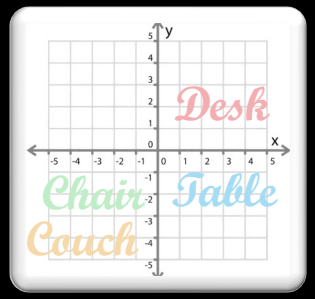
Retail Example

Similarity Search

Product
Embeddings

stomer
oduct
tabase

Azure
OpenAI
Embeddings



Embeddings Space

Couch

Relevant sort
Price : **Lower first**

Similarity Search

Kichine $2,300
Signature series

Olova $3,800
Radiant series

Kinsu $250
World series

Squarrio $287
8 pcs set

Marketing Example

Digital Creative Assistant – Dalle-2

Synthetic Brand Ambassador (GDPR-safe)

Input

Generate a white female

Generate

Output

Generate images, concepts and ideas

Input

a green c

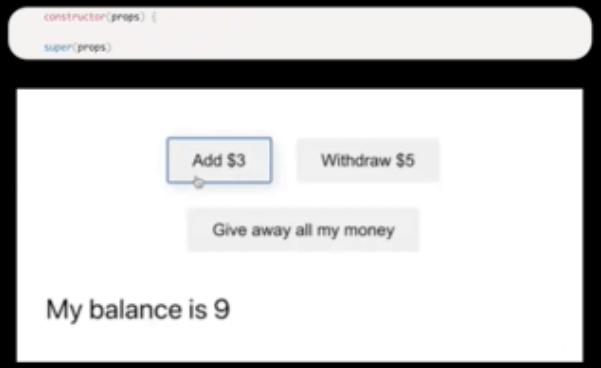
Output

No code Web and App Development

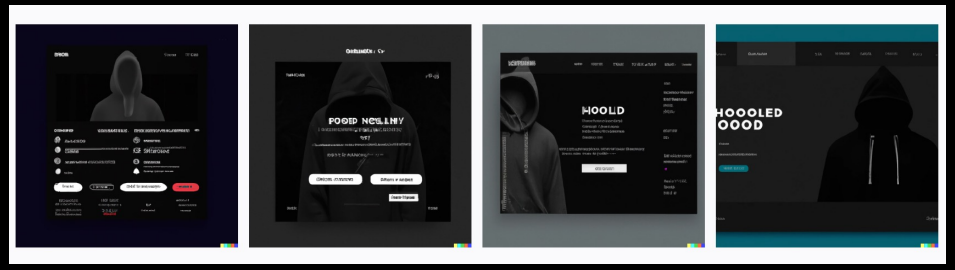
Input

a button that says "add $3" and a button that says "withdraw $5", and a button that says "Give away all my money", then show me my balance

Output



Produce rough Layouts



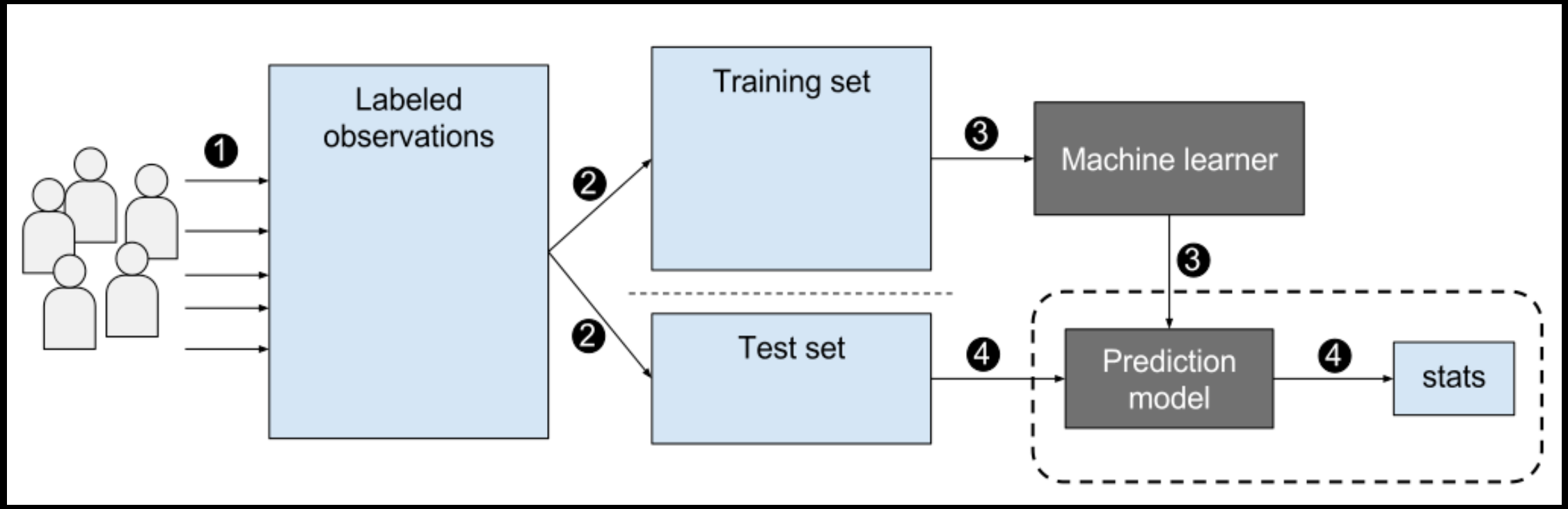
Agenda

Big Data & AI

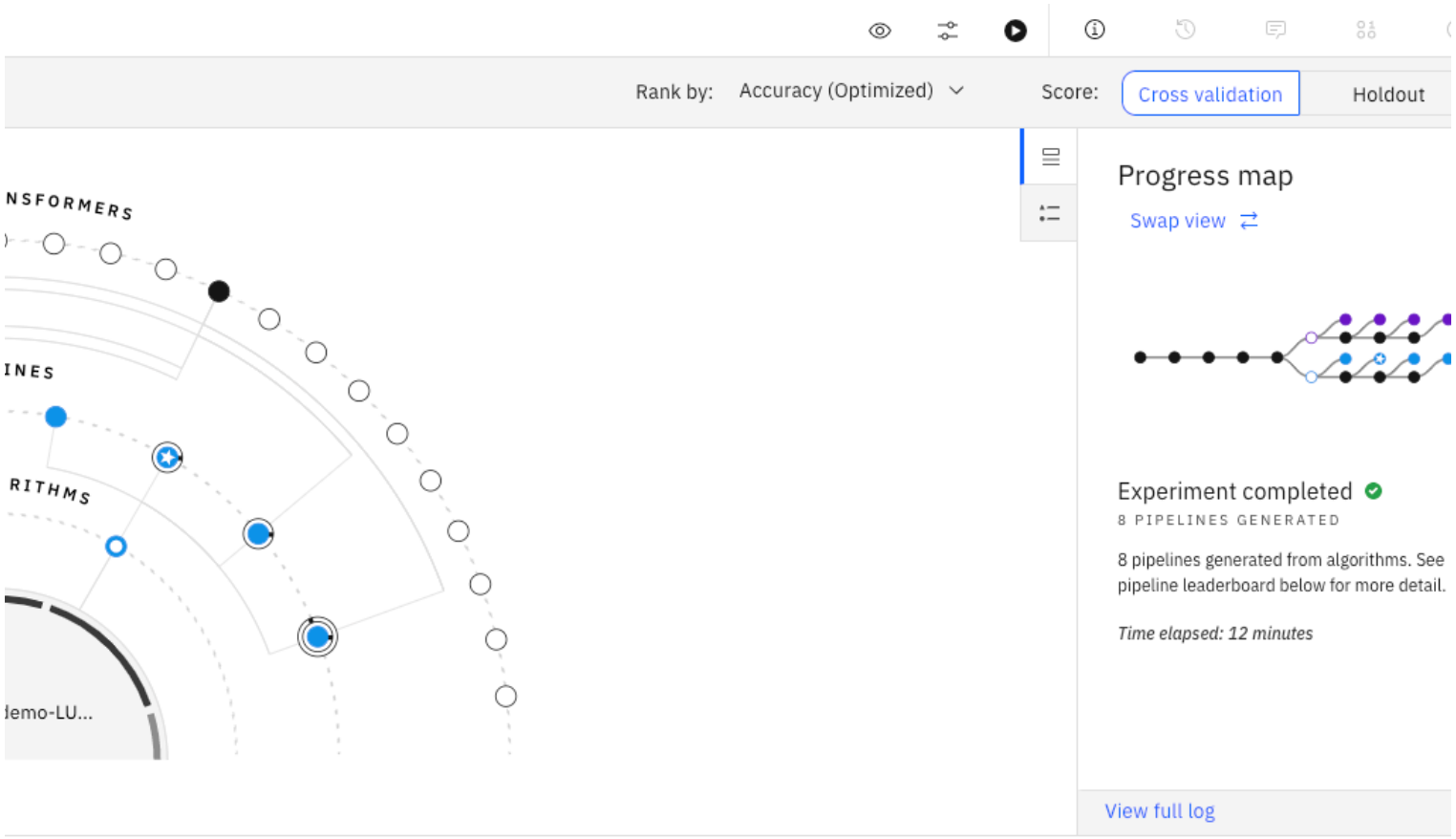
Generative AI (ChatGPT)

Demo

Supervised Machine Learning



Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events



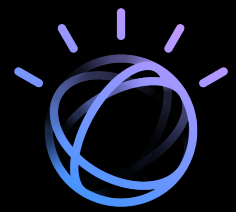
Accuracy (Optimized)	Enhancements	Build time	
0.975	HPO-1	00:01:00	Save as
0.974	HPO-1 FE	00:05:31	

Supervised Machine Learning Demo

Closing Thoughts

Industry Principles for Trust and Transparency

https://www.ibm.com/downloads/cas/J45XZOAR

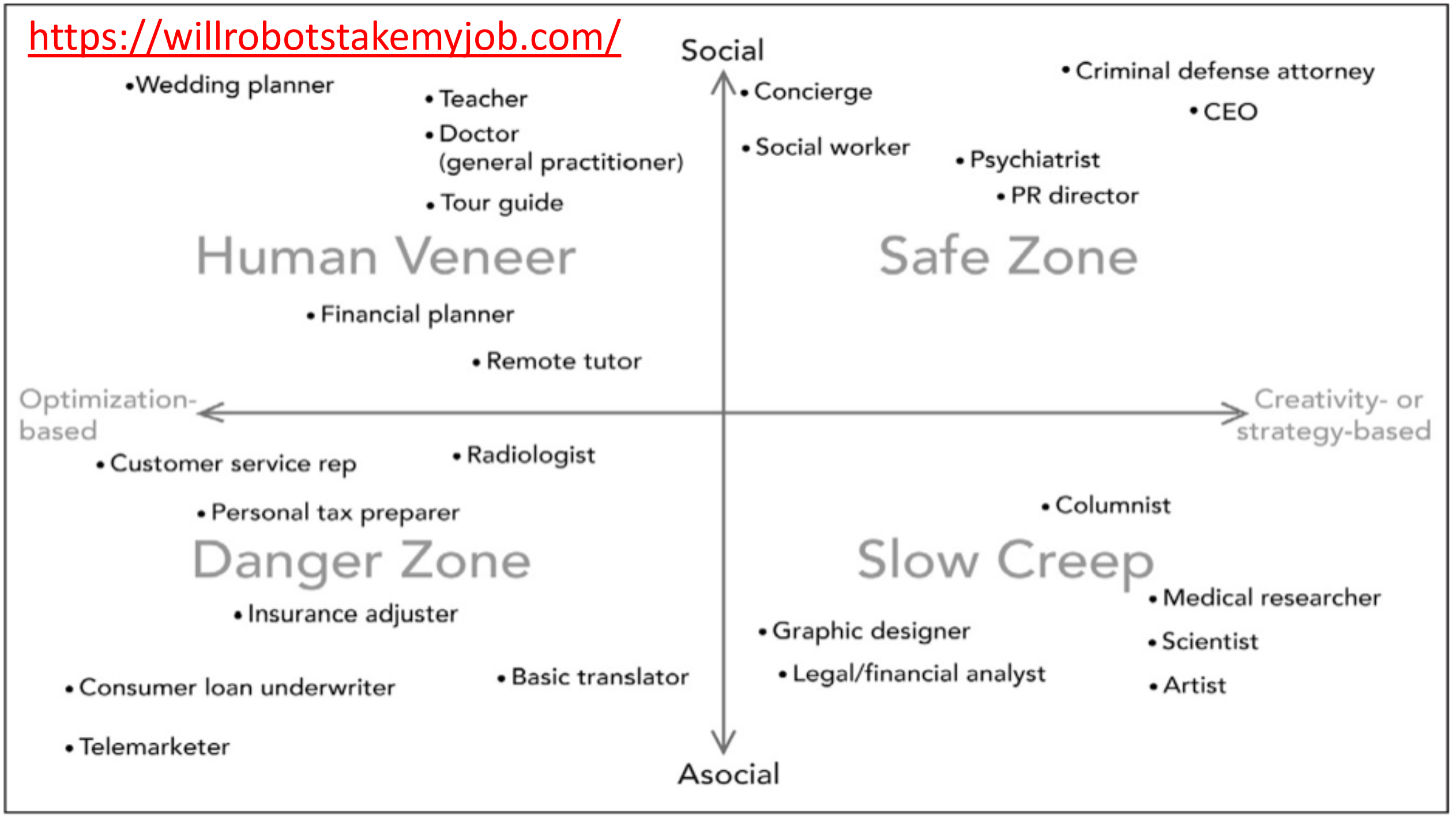


1. The purpose of AI is to augment human intelligence

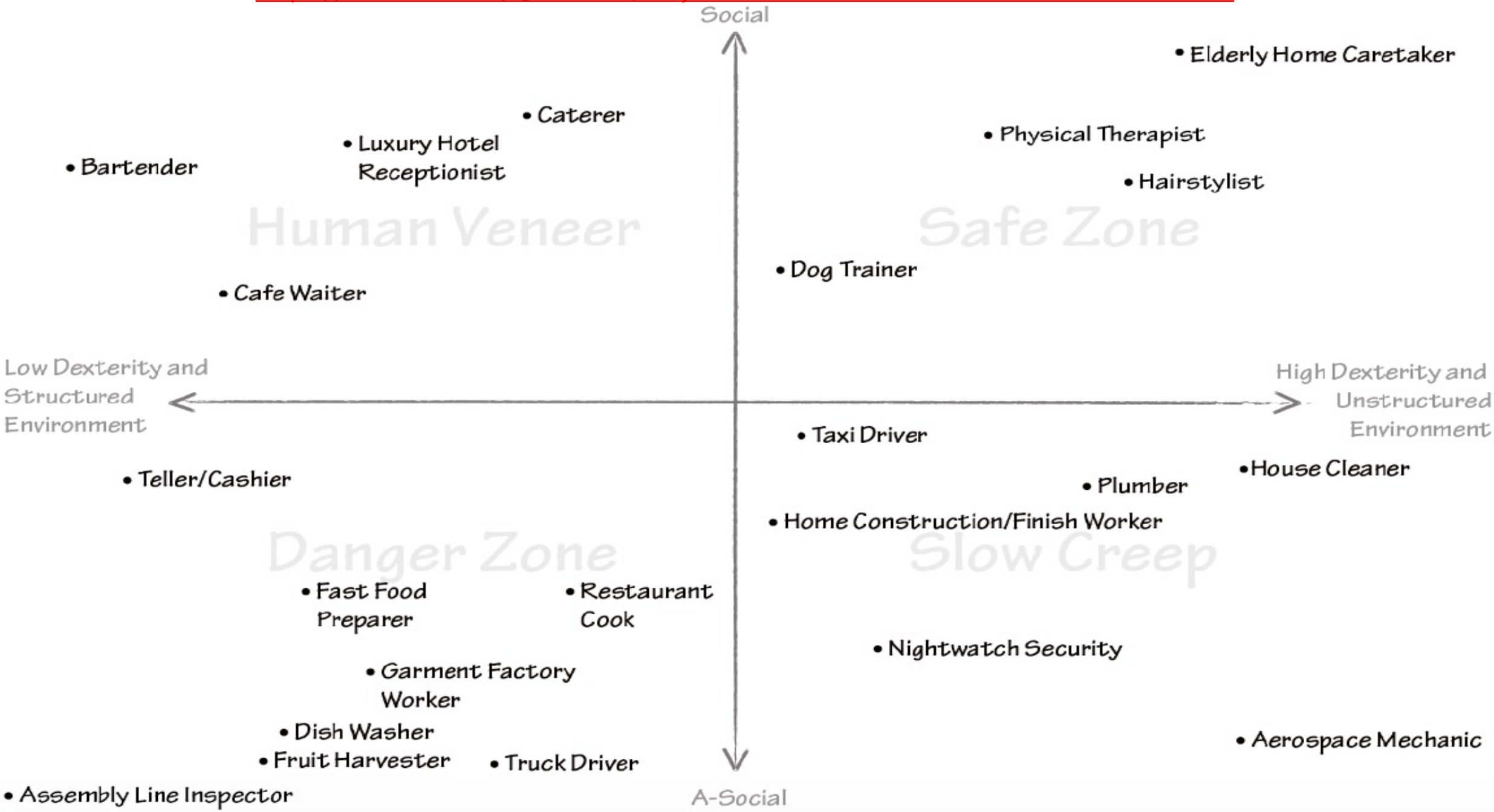
2. Data and insights belong to their creator

3. AI systems must be transparent and explainable

https://willrobotstakemyjob.com/



https://medium.com/@kaifulee/10-jobs-that-are-safe-in-an-ai-world-ec4c45523f4f



Takeaways AI lecture

Citizen AI:

Develop critical thinking skills, have your voice heard

https://www.accenture.com/cz-en/insight-explainable-citizen-ai

Big Data Advantage:

When giving away your personal data, choose wisely

https://www.slideshare.net/RobertoVll/ai-and-big-data-for-business-and-people-advantage

Intelligent Automation:

Educate yourself on AI, thrive in your profession

https://1.dam.s81c.com/m/3de136737e51fb20/original/IBM-Automation-Whitepaper_Final.pdf

Q&A

Ivan Portilla

Portilla@gmail.com