

Generative AI Playbook by PromptxAI and ChatStart

Build ChatGPT and other Generative AI powered startup. Fast!

PromptxAI and ChatStart Team

2022 to present, PromptxAI and ChatStart Team. All rights reserved.

Table of Contents

I. Generative AI Playbook by PromptxAI	5
I.1 Playbook with popular examples	5
I.2 Extended playbook with a generic framework	6
Content generation	7
I.3 Trends impacting future of work	7
I.4 Trend 1: Generative AI	7
I.5 Trend 2: Prompt Engineering	9
I.6 Trends 3: Creator Hubs	9
2. ChatGPT	11
2.1 Historical Fact Question Answer	11
Attribute related to an entity	11
Citing source of historical fact	11
List of historical facts	12
Create dataset of historical facts	12
2.2 Generating Diagrams From World Knowledge	14
Visualizing movie scripts	14
Generating visual summaries of books	15
Automating diagram generation	16
2.3 Natural Language Analytics and Visualization	17
2.4 Tallest Buildings Dataset Visualization	19
3. Media AI	21
3.1 Cinematic Alien Art Generation	21
Cinematic alien geometries	21
Organic alien objects in super detailed 3D	23
Mesmerizing alien landscapes	25
3.2 Interview with Stability AI CEO Emad Mushtaq	29
Transcript	30
4. Hugging Face	62
4.1 5 Minutes Generative AI Developer Setup	62
5. Open AI	65
5.1 OpenAI CEO Sam Altman on AI for the Next Era	65
Introduction	65
Data Flywheel Process	68
Business Opportunities	69
Future of AI	71

Target Markets	72
AI and the Metaverse	74
Video Transcript Continued	75
6. Models	79
6.1 Large Language Models	79
Why Large Language Models Matter	79
List of Large Language Models	79
6.2 DALL-E Mini	81
Model Summary	81
Model Resources	81
Model Details	81
6.3 LaMDA	83
Model Summary	83
Model Resources	83
Model Details	83
6.4 LLaMA	85
Model Summary	85
Model Resources	85
Model Details	85
6.5 BERT base model (uncased)	87
Model Summary	87
Model Resources	87
Model Details	87
6.6 Generative Pre-trained Transformer 3 (GPT-3)	89
Model Summary	89
Model Resources	89
Model Details	89
6.7 Flan 20B	91
Model Summary	91
Model Resources	91
Model Details	91
7. Tags	92
7.1 Analytics Function	92
7.2 Apache	92
7.3 ChatGPT	92
7.4 ChatGPT Model	92
7.5 Consulting Industry	92
7.6 Creative Function	92

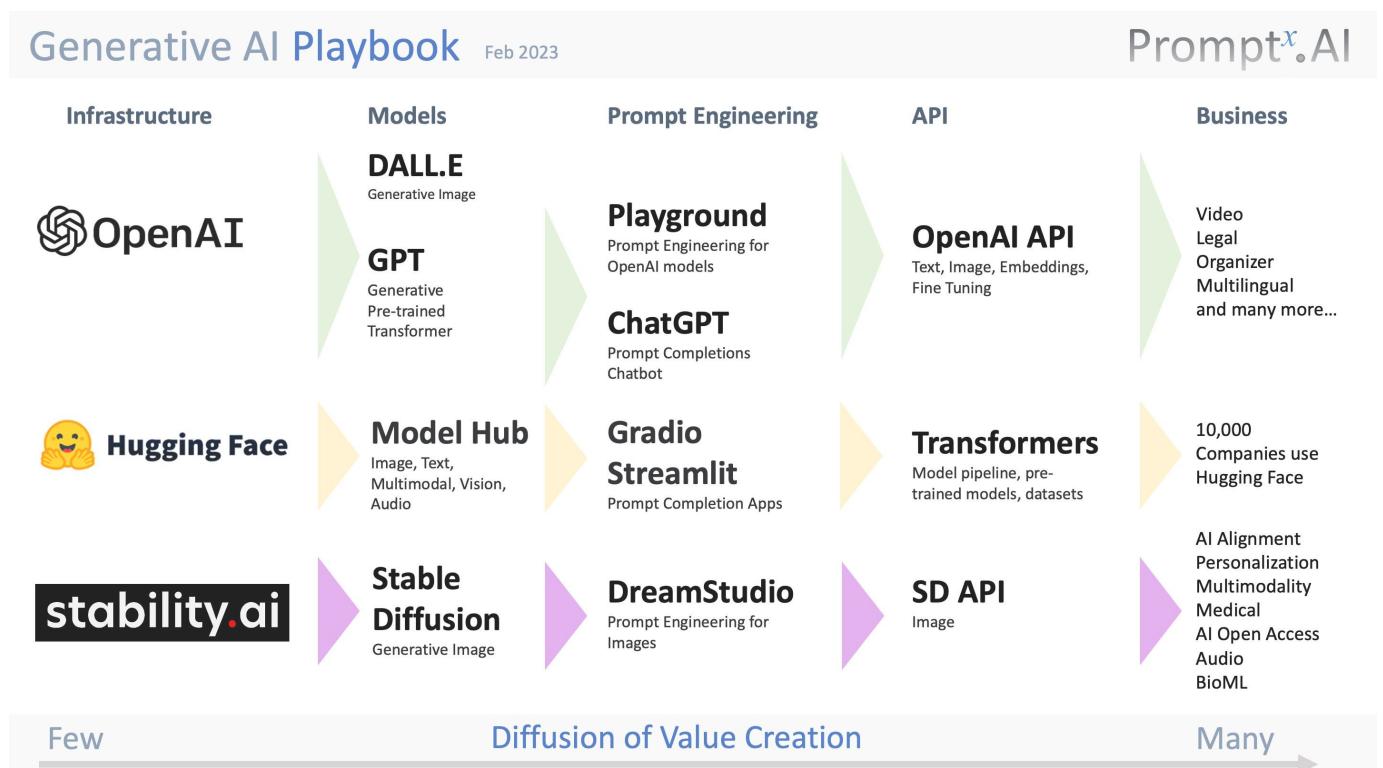
7.7 Dataset Generation	92
7.8 Deep learning	92
7.9 Diagram Generation	93
7.10 Education Industry	93
7.11 English language	93
7.12 Film Industry	93
7.13 Flan	93
7.14 Generative AI	93
7.15 Generative pre-training	93
7.16 Google AI	93
7.17 Hugging Face	93
7.18 Image Generation	94
7.19 Image generation	94
7.20 Instruction tuning	94
7.21 Knowledge Management	94
7.22 Language Models for Dialog Applications	94
7.23 Language model	94
7.24 MLM	94
7.25 Masked language modeling	94
7.26 Midjourney Model	94
7.27 Natural language processing (NLP)	94
7.28 OpenAI	95
7.29 Research Function	95
7.30 Stability AI	95
7.31 T5	95
7.32 Transformer	95
7.33 Transformers model	95
7.34 UL2	95
7.35 transformer architecture	95

1. Generative AI Playbook by PromptxAI

Generative AI is disrupting many industries and significantly reshaping future of work for many creative and knowledge-driven functions. PromptxAI is on a mission to help businesses and individuals leverage the power of Generative AI to accelerate their growth and innovation. We are building a Generative AI Playbook to help you get started with Generative AI. You can access the Generative AI Playbook on this website in various ways including learning creative prompt engineering [techniques](#) to make the most of Generative AI applications like [ChatGPT](#), Midjourney, DALL-E, [Stable Diffusion](#), D-ID, and more. You can also learn about the [trends](#) impacting future of work and how to leverage these as ideas for your own business or career. We are also sharing accelerator [notebooks](#) and code [libraries](#) to help you get started with Generative AI.

1.1 Playbook with popular examples

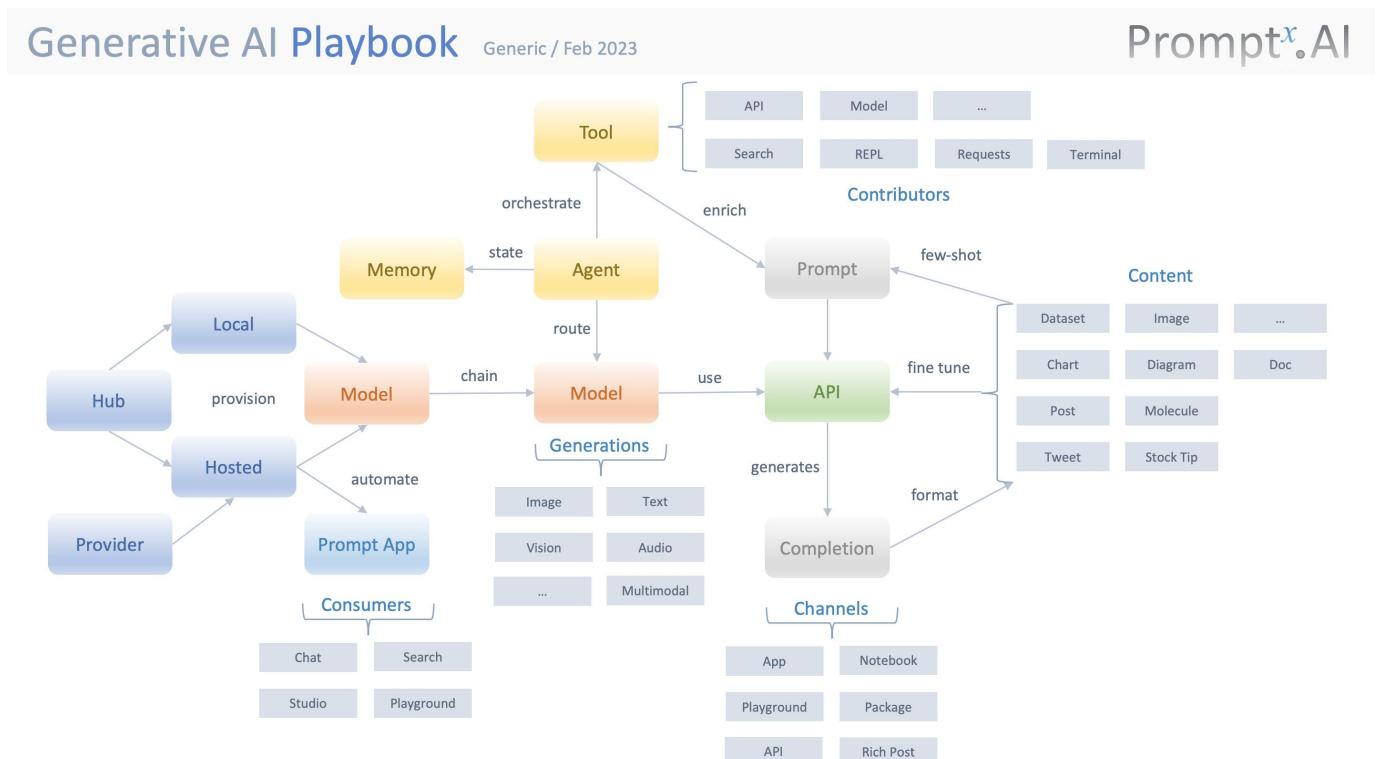
Here is a version of the playbook with most popular examples.



1.2 Extended playbook with a generic framework

Here is a generic version of the playbook which extends the Generative AI landscape with Agents, Tools, and Content. This playbook extends and uses [LangChain](#) as a generic framework for working with LLMs.

- 1. Prompts:** This includes prompt management, prompt optimization, and prompt serialization.
- 2. Models:** This includes LLMs and pre-trained deep learning transformer models.
- 3. Content:** This includes content which can be used to enhance a prompt with few-shot learning, content which can fine-tune an LLM, or content which can be enhanced or generated by LLM completions.
- 4. Tools:** Language models are often more powerful when interacting with other sources of knowledge or computation. This can include Python REPLs, embeddings, search engines, and more.
- 5. Chains:** Chains go beyond just a single LLM call, and are sequences of calls (whether to an LLM or a different utility).
- 6. Agents:** Agents involve an LLM making decisions about which Actions to take, taking that Action, seeing an Observation, and repeating that until done.
- 7. Memory:** Memory is the concept of persisting state between calls of a chain/agent.
- 8. Infrastructure:** This includes hubs like Hugging Face, GitHub, and Kaggle. It also includes cloud providers like AWS, GCP, and Azure. First movers like OpenAI and Stability AI are hosted providers of LLMs. Another emerging category is specialized multi-cloud orchestrators for LLM training, optimization, and inference.



Content generation

You can explore the generic playbook on this website. Let's start with far right part of the playbook, Content. An example of prompt-model-completion sequence generating content in the format of a diagram is demonstrated in [Generating Diagrams from World Knowledge](#) section. You can also explore [Low Code Blog Authoring Automation](#) section to learn about how to use PromptxAI API to automate blog authoring. The post on [Natural Language Analytics and Visualization](#) demonstrates charts and graphs. The [OpenAI CEO Sam Altman on AI for Next Era](#) interview demonstrates creating topic maps and summaries as content. You can generate datasets like [Historical Fact Question Answer](#) does for ancient civilizations.

1.3 Trends impacting future of work

There are several trends impacting future of work. We are concerned about impact on the knowledge worker and startup community. Here are some trends we believe are at the heart of the workplace disruption as explained by ChatGPT Transformer model (leading example of trend 1) generated completions or responses based on prompts or inputs we have provided (example of trend 2). We have coined a shared name for GitHub, Kaggle, and Huggingface as trend 3, calling them Creator Hubs.

1.4 Trend 1: Generative AI

ChatGPT, DALL-E, Midjourney, Stable Diffusion, D-ID are popular applications of Generative AI. While ChatGPT generates human-like responses as a chatbot, DALL-E, Midjourney, and Stable Diffusion can generate images from text prompts. D-ID generates videos. Behind the scenes these applications use Transformer models such as GPT-3, also known as Large Language Models.

ChatGPT Prompt

Explain Generative AI with well known examples. Cite sources.

ChatGPT Completion

Generative AI is a type of artificial intelligence that focuses on creating new data from existing data. It is used to generate new data that is similar to existing data, but not identical. Generative AI can be used to create new images, music, text, and videos.

One of the most well-known examples of generative AI is Google's DeepDream. DeepDream is a computer vision program that uses a deep neural network to generate new images based on existing images. It can be used to create surreal and abstract images from existing photographs.

Another example of generative AI is OpenAI's GPT-3. GPT-3 is a natural language processing system that can generate text based on existing text. It can be used to generate new stories, articles, and other written content.

Finally, generative AI can also be used to create new music. Google's Magenta project uses generative AI to create new music based on existing music. It can be used to create new compositions or remix existing songs.

Sources:

- <https://www.analyticsinsight.net/what-is-generative-ai/>
- <https://www.forbes.com/sites/cognitiveworld/2020/07/14/what-is-generative-ai-and-how-it-works/#3f9f9f9f3f2f>
- <https://www.towardsdatascience.com/generative-ai-what-it-is-and-how-it-works-f9f9f9f3f2f>

1.5 Trend 2: Prompt Engineering

Prompt Engineering is right now the buzz on social media. It is interesting how Sam Altman addresses future of Prompt Engineering in the interview (25:09) where he thinks it will evolve to people just interacting with AGI in a natural way. However, Reid Hoffman makes a counterpoint how an artist can articulate much better prompts and get more out of Transformer Models than a casual user.

ChatGPT Prompt

Explain Prompt Engineering as it applies to models like ChatGPT and Midjourney

ChatGPT Completion

Prompt engineering is a technique used to improve the performance of natural language processing (NLP) models such as ChatGPT and Midjourney. It involves carefully crafting the input to the model to ensure that it produces the desired output. This can involve changing the wording of the input, adding additional context, or providing additional information. By doing this, the model can better understand the user's intent and provide more accurate and relevant responses. Prompt engineering can also be used to improve the model's ability to generate natural-sounding responses.

1.6 Trends 3: Creator Hubs

To understand the gap between successful entrepreneurs or startup teams and the typical knowledge worker at an established enterprise, one can compare their tooling, work habits, and skills. Alternatively, one can look at the rapid clip of advances made by open source and citizen data science communities versus those of market leaders in the same space. The acquisitions of Kaggle by Google in 2017 (financial terms undisclosed), GitHub by Microsoft in 2018 (for \$7.5 billion), Streamlit by Snowflake in 2021 (for \$800 million), the \$2 billion valuation of Huggingface, and the projected \$20 billion valuation of OpenAI at the end of 2022 demonstrate the impact of these new market makers on the future of work and the growth of a community of knowledge workers learning new skills. Hugging Face offers Streamlit Spaces for hosting apps built on Streamlit. OpenAI GPT-2 model is available on Hugging Face Models repository (as on Feb'23 hosting more than 133K pre-trained models).

ChatGPT Prompt

Briefly explain what is GitHub, Kaggle, and Huggingface.

What does success of GitHub, Huggingface, and Kaggle have in common?

Explain as an expert in the subject using a few paragraphs in narrative style.

ChatGPT Completion

GitHub is a platform for version control and collaboration on software projects. It allows developers to share and work on code together, and offers a variety of tools and features to make it easier for developers to collaborate and manage their projects.

Kaggle is a platform for data science competitions and collaboration on data science projects. It offers a variety of datasets, as well as tools for data exploration, visualization, and machine learning model development. Kaggle is popular among data scientists and analysts who want to learn from and work with others in the field.

Huggingface is a platform for working with natural language processing (NLP) models. It provides access to a large collection of pre-trained models and tools for working with them, and is popular among researchers and developers working on NLP projects.

The success of GitHub, Huggingface, and Kaggle can be attributed to a few key factors.

First, all three platforms offer valuable resources and tools that are useful to developers and data scientists. GitHub is a popular platform for version control and collaboration on software projects, and it offers a variety of tools and features that make it easier for developers to work together on code. Huggingface is a platform for working with natural language processing (NLP) models, and it provides access to a large collection of pre-trained models and tools for working with them. Kaggle is a platform for data science competitions and collaboration on data science projects, and it offers a variety of datasets, as well as tools for data exploration, visualization, and machine learning model development.

Another factor contributing to the success of these platforms is their focus on community and collaboration. All three platforms provide a way for users to share their work with others and collaborate on projects, which can be a valuable resource for developers and data scientists. This helps to create a sense of community and encourages users to contribute their knowledge and skills.

Finally, the success of these platforms can also be attributed to their user-friendly interfaces and ease of use. GitHub, Huggingface, and Kaggle all have well-designed websites that are easy to navigate and use, which helps to make them accessible to a wide range of users.

Overall, the success of GitHub, Huggingface, and Kaggle can be attributed to the valuable resources and tools they offer, their focus on community and collaboration, and their user-friendly interfaces. These factors have helped to make these platforms popular among developers and data scientists, and have contributed to their success in the tech industry.

2. ChatGPT

 Research Function  Education Industry  ChatGPT Model  Dataset Generation

2.1 Historical Fact Question Answer

You can ask ChatGPT any questions about historical facts and it will respond with the correct answer. ChatGPT is trained on a knowledge corpus of historical facts which includes Wikipedia and Encyclopedia Britannica among other sources.

Guide

Ask question related to a historical fact about any entity like a famous personality, book, movie, product, or place and related attributes like location, age, distance, size, or weight.

Attribute related to an entity

Here is an example of a simple prompt-completion sequence using ChatGPT. The human is asking a factual question. The model is asked to complete the question with the correct answer.

ChatGPT Prompt

Where is the great pyramid located?

ChatGPT responds with a completion which is based on the knowledge corpus used for training the model.

ChatGPT Completion

The Great Pyramid is located in Giza, Egypt.

Citing source of historical fact

You could get more information from ChatGPT like citing the source of the answer.

ChatGPT Prompt

Where is the great pyramid located? Cite the source of answer.

ChatGPT responds with a completion along with source of the answer.

ChatGPT Completion

The Great Pyramid is located in Giza, Egypt. Source: <https://www.britannica.com/topic/Great-Pyramid-of-Giza>

List of historical facts

You could also get structured data as completion by asking for the type of structure to return as completion.

ChatGPT Prompt

```
List the 10 most popular pyramids and their locations.
```

ChatGPT responds with a completion which is a list of popular pyramids and their locations. Note that not all of them are located in Egypt.

ChatGPT Completion

1. Great Pyramid of Giza, Egypt
2. Khafre's Pyramid, Egypt
3. Menkaure's Pyramid, Egypt
4. Djoser's Pyramid, Egypt
5. Pyramid of the Sun, Teotihuacán, Mexico
6. Pyramid of the Moon, Teotihuacán, Mexico
7. El Castillo, Chichen Itza, Mexico
8. Temple of the Sun, Peru
9. Temple of the Inscriptions, Palenque, Mexico
10. Bent Pyramid, Egypt

Create dataset of historical facts

You can create a dataset of historical facts related entities and their attributes.

Prompt

```
Only respond using fenced code block with markdown which I can copy-paste  
to an editor. #(1)!
```

```
Tell me about the entity Mayan Civilisation. #(2)!
```

```
Create a table of few attributes of this entity in attribute name
```

and value columns. Then create a table of 10 related entities with these attributes names as columns and values for each of the entities as rows.

1. ✎ You can leave this out if you want to get the completion in a plain text format. In this case the table will be rendered in HTML.
- 2.💡 You can replace Mayan Civilization with any other historical entity.

ChatGPT responds with a completion which is a table of attributes and related entities as markdown in a fenced code block.

✍ ChatGPT Completion

Mayan Civilisation

Attribute Name	Attribute Value
Location	Central America, present-day Mexico, Guatemala, Belize, Honduras, El Salvador and parts of Nicaragua
Time Period	2000 BC to 16 th century AD
Religion	Polytheistic, worshipped gods related to natural phenomena such as the sun, the moon, and agriculture
Writing System	Hieroglyphic writing system, one of the few writing systems of the pre-Columbian Americas
Architecture	Elaborate stone pyramids, plazas, palaces, and elaborate underground structures such as tombs and cisterns
Mathematics and Astronomy	Highly advanced knowledge of mathematics and astronomy, including the use of zero as a place-holder and the calculation of a solar year to within five minutes
Art	Highly developed art style, including elaborate murals, carvings, pottery, and weaving
Collapse	The Mayan civilization declined and eventually collapsed in the 9 th to 10 th centuries, with reasons including environmental degradation, war, and societal problems

Related Entities

Entity	Location	Time Period	Religion	Writing System
Olmecs	Mexico	1400 BC - 400 BC	Polytheistic	Undeciphered script
Teotihuacan	Mexico	300 BC - 700 AD	Polytheistic	None
Aztecs	Mexico	13 th to 16 th century AD	Polytheistic	Nahuatl writing system
Inca Empire	South America	1438 AD to 1533 AD	Polytheistic	Quipu system of knots and colors
Toltecs	Mexico	900 AD to 1168 AD	Polytheistic	None
Zapotecs	Mexico	600 BC - 1521 AD	Polytheistic	Zapotec writing system
Mixtecs	Mexico	300 AD - 1521 AD	Polytheistic	Mixtec writing system
Maya of Yucatán	Mexico	600 AD to 1697 AD	Polytheistic	Hieroglyphic writing system
Chichimeca	Mexico	900 AD to 1500 AD	Polytheistic	None
Purépecha	Mexico	1000 AD to 1530 AD	Polytheistic	None

2.2 Generating Diagrams From World Knowledge

Mermaid.js is a JavaScript library that can be used to generate diagrams from text using a simple, human-readable syntax. You can ask ChatGPT to generate code for a diagram using the Mermaid.js syntax.

i Guide

Describe what the diagram should illustrate and what kind of diagram to generate. Instruct ChatGPT to generate Mermaid.js code for the diagram.

Visualizing movie scripts

In the following prompt we are expecting ChatGPT to decipher the contents of the diagram based its knowledge of the movie Inception. We do not have to explicitly describe any of the characters, locations, or events in the movie.

ChatGPT Prompt

Generate Mermaid.js code for a context diagram that shows the main characters and their relationships in the movie Inception. # (1)!
Include their interactions with external entities such as locations and events, using the →| Relationship | syntax to indicate the relationships between the different entities, using subgraphs for characters, locations, and events.

- I. 💡 Replace `movie Inception` with name of any movie, book, or play you want to generate a context diagram for.

The completion from ChatGPT is Mermaid.js code that can be used to generate a context diagram like below.

```
graph LR;
    subgraph "Characters"
        Cobb(Cobb)
        Mal(Mal)
        Arthur(Arthur)
        Ariadne(Ariadne)
        Eames(Eames)
        Yusuf(Yusuf)
        Saito(Saito)
    end
    subgraph "Locations"
        Paris(Paris)
        Tokyo(Tokyo)
        USA(USA)
    end
    subgraph "Events"
        Inception(Inception)
    end
    Cobb --> Mal
    Cobb --> Arthur
    Cobb --> Ariadne
    Cobb --> Eames
    Cobb --> Yusuf
    Cobb --> Saito
    Paris --> Inception
    Tokyo --> Inception
    USA --> Inception
    Cobb -->|Married To| Mal
    Cobb -->|Works With| Arthur
```

```

Cobb -->|Works With| Ariadne
Cobb -->|Works With| Eames
Cobb -->|Works With| Yusuf
Cobb -->|Works For| Saito
Cobb -->|Located In| Paris
Cobb -->|Located In| Tokyo
Cobb -->|Located In| USA
Mal -->|Located In| Paris
Arthur -->|Located In| Paris
Ariadne -->|Located In| Paris
Eames -->|Located In| Paris
Yusuf -->|Located In| Paris
Saito -->|Located In| Paris

```

Generating visual summaries of books

Let us try a variation of the prompt. This time we are expecting ChatGPT to decipher the contents of the diagram based its knowledge of the book The Unicorn Project.

OpenAI GPT-3 Prompt

```

Generate Mermaid.js code for a context diagram that shows the main entities and their relationships in the book The Unicorn Project.
Include the interactions between entities using the ->| Relationship | syntax to indicate the relationships between the different entities, using subgraphs clustering entities of similar class.

```

Here is the context diagram generated for the book.

```

graph LR;
subgraph The Unicorn Project
    DevOpsTeam[DevOps Team] -->|manages| DevOpsToolchain[DevOps Toolchain]
    DevOpsTeam -->|uses| DevOpsPipeline[DevOps Pipeline]
    DevOpsTeam -->|uses| CIPlatform[CI Platform]
    DevOpsTeam -->|uses| PerformanceTesting[Performance Testing]
    DevOpsTeam -->|uses| SecurityTesting[Security Testing]
    DevOpsTeam -->|uses| AutomatedTesting[Automated Testing]
    DevOpsTeam -->|uses| Monitoring[Monitoring]
end

subgraph The Business
    BusinessOwner[Business Owner] -->|oversees| ProductOwner[Product Owner]
    ProductOwner -->|manages| BusinessTeam[Business Team]
end

subgraph The Engineering Team
    EngineeringTeam[Engineering Team] -->|manages| EngineeringToolchain[Engineering Toolchain]
    EngineeringTeam -->|uses| VersionControl[Version Control]
    EngineeringTeam -->|uses| TDD[TDD]
    EngineeringTeam -->|uses| Refactoring[Refactoring]
    EngineeringTeam -->|uses| PairProgramming[Pair Programming]
    EngineeringTeam -->|uses| CodeReview[Code Review]
end

DevOpsToolchain --> CIPlatform
DevOpsToolchain --> PerformanceTesting
DevOpsToolchain --> SecurityTesting
DevOpsToolchain --> AutomatedTesting
DevOpsToolchain --> Monitoring

EngineeringToolchain --> VersionControl
EngineeringToolchain --> TDD
EngineeringToolchain --> Refactoring
EngineeringToolchain --> PairProgramming
EngineeringToolchain --> CodeReview

CIPlatform --> DevOpsPipeline
DevOpsPipeline --> BusinessTeam
BusinessTeam --> ProductOwner

```

Automating diagram generation

You can write automation to generate such diagrams from world knowledge using OpenAI API within few lines of code. In the following code we import the OpenAI API, initialize the API key, and generate the completion in less than five lines of code.

OpenAI API

```
import openai
openai.api_key = os.getenv("OPENAI_KEY")
openai_model = 'text-davinci-003'
completion = openai.Completion.create(
    prompt=prompt,
    model=openai_model,
    temperature=0.6,
    max_tokens=500)
mermaid_code = completion.choices[0].text
```

 **Analytics Function**  **ChatGPT**

2.3 Natural Language Analytics and Visualization

We can generate datasets using ChatGPT prompts. Then we can combine use these datasets for visualization and analytics.

ChatGPT Prompt

```
What are the house price trends over last 10 years in Sunnyvale, CA?
Only provide historically factual data and cite source of data.
In response create a markdown table starting # (1) !
with "```" markdown and closing with "```".
```

- I.  Note that in the prompt "```" needs to be mentioned without enclosing quotes. We have used the enclosing quotes to ensure it renders as-is in our code block. You can also skip this part of the prompt if you do not want to copy-paste the markdown code for the table.

This generates a table of data and citing source of the data. Note that as you run this query a few times, the source of data and the values may vary. Asking for the source of the data along with the data ensures that the data is historically factual and not made up. You could even perform a sanity check by comparing the data with the source of data.

ChatGPT Completion

Year	Median House Price in Sunnyvale, CA
2011	\$756,000
2012	\$833,750
2013	\$940,000
2014	\$1,148,000
2015	\$1,360,000
2016	\$1,450,000
2017	\$1,555,000
2018	\$1,710,000
2019	\$1,730,000
2020	\$1,900,000

Source: Zillow Home Value Index. (n.d.). Sunnyvale home values. Zillow. <https://www.zillow.com/sunnyvale-ca/home-values/>

Next we instruct ChatGPT within the same chat session to generate code for [vegalite charts](#). Vega-Lite is a high-level grammar of interactive graphics. It provides a concise, declarative JSON syntax to create an expressive range of visualizations for data analysis and presentation.

ChatGPT Prompt

```
Visualize this table data by generating code for vegalite charts
using bar charts.
```

Here is the visualization for the code that ChatGPT generates.

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "data": {
    "values": [
      {"Year": 2011, "Median House Price in Sunnyvale, CA": 756000},
      {"Year": 2012, "Median House Price in Sunnyvale, CA": 833750},
      {"Year": 2013, "Median House Price in Sunnyvale, CA": 940000},
      {"Year": 2014, "Median House Price in Sunnyvale, CA": 1148000},
      {"Year": 2015, "Median House Price in Sunnyvale, CA": 1360000},
      {"Year": 2016, "Median House Price in Sunnyvale, CA": 1450000},
      {"Year": 2017, "Median House Price in Sunnyvale, CA": 1555000},
      {"Year": 2018, "Median House Price in Sunnyvale, CA": 1710000},
      {"Year": 2019, "Median House Price in Sunnyvale, CA": 1730000},
      {"Year": 2020, "Median House Price in Sunnyvale, CA": 1900000}
    ]
  },
  "mark": {"type": "bar", "tooltip": true},
  "encoding": {
    "y": {"field": "Year", "type": "nominal"},
    "x": {"field": "Median House Price in Sunnyvale, CA", "type": "quantitative", "format": "$,.0f"}
  },
  "color": {"field": "Median House Price in Sunnyvale, CA", "type": "quantitative", "scale": {"scheme": "bluepurple"}},
  "config": {"axis": {"labelFontSize": 14, "titleFontSize": 16, "labelFontWeight": "bold"}}
}
```

As you can see vegalite is an awesome library for visualizing data and the chart is interactive. You can hover over the bars to see the values. ChatGPT takes care of generating styled charts with gradients, colors, and legends, without any need for us to provide these instructions as part of the prompt.

In summary, you can now use natural language to query world knowledge learnt by a large language model like ChatGPT and generate visualizations and analytics. No databases, no SQL, no programming, no coding, no data engineering, no JavaScript knowledge required!

2.4 Tallest Buildings Dataset Visualization

Let us pull a dataset of tallest buildings.

ChatGPT Prompt

Create a table of tallest buildings in the world with various attributes like height, year built, location, and others. Draw the table in markdown enclosing in "```".

Generates this table with extra attributes including the ones specified in the prompt.

Rank	Building Name	Height (m)	Height (ft)	Year Completed	City	Country	Primary Use
1	Burj Khalifa	828	2,716	2010	Dubai	United Arab Emirates	Mixed Use
2	Shanghai Tower	632	2,073	2015	Shanghai	China	Office
3	Abraj Al Bait Clock Tower	601	1,971	2012	Mecca	Saudi Arabia	Mixed Use
4	Ping An Finance Center	599	1,965	2017	Shenzhen	China	Office
5	Lotte World Tower	555	1,819	2017	Seoul	South Korea	Mixed Use
6	One World Trade Center	541	1,776	2014	New York City	United States	Office
7	Guangzhou CTF Finance Centre	530	1,739	2016	Guangzhou	China	Hotel/ Residential/ Office
8	Tianjin CTF Finance Centre	530	1,739	2019	Tianjin	China	Hotel/Office
9	CITIC Tower	528	1,732	2018	Beijing	China	Office
10	TAIPEI 101	508	1,667	2004	Taipei	Taiwan	Office

Now we define a prompt for visualizing the table using vegalite charts.

ChatGPT Prompt

Visualize this table choosing appropriate vegalite chart and generate the code for the visualization.

The generated code handles mixed data types intelligently and is able to identify appropriate chart design based on given attributes and types.

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "data": {
    "values": [
      {"Building Name": "Burj Khalifa", "Height (m)": 828, "Height (ft)": 2716, "Year Completed": 2010, "City": "Dubai", "Country": "United Arab Emirates", "Primary Use": "Mixed Use"}, {"Building Name": "Shanghai Tower", "Height (m)": 632, "Height (ft)": 2073, "Year Completed": 2015, "City": "Shanghai", "Country": "China", "Primary Use": "Office"}, {"Building Name": "Abraj Al Bait Clock Tower", "Height (m)": 601, "Height (ft)": 1971, "Year Completed": 2012, "City": "Mecca", "Country": "Saudi Arabia", "Primary Use": "Mixed Use"}, {"Building Name": "Ping An Finance Center", "Height (m)": 599, "Height (ft)": 1965, "Year Completed": 2017, "City": "Shenzhen", "Country": "China", "Primary Use": "Office"}, {"Building Name": "Lotte World Tower", "Height (m)": 555, "Height (ft)": 1819, "Year Completed": 2017, "City": "Seoul", "Country": "South Korea", "Primary Use": "Mixed Use"}, {"Building Name": "One World Trade Center", "Height (m)": 541, "Height (ft)": 1776, "Year Completed": 2014, "City": "New York City", "Country": "United States", "Primary Use": "Office"}, {"Building Name": "Guangzhou CTF Finance Centre", "Height (m)": 530, "Height (ft)": 1739, "Year Completed": 2016, "City": "Guangzhou", "Country": "China", "Primary Use": "Hotel/Residential/Office"}, {"Building Name": "Tianjin CTF Finance Centre", "Height (m)": 530, "Height (ft)": 1739, "Year Completed": 2019, "City": "Tianjin", "Country": "China", "Primary Use": "Hotel/Office"}, {"Building Name": "CITIC Tower", "Height (m)": 528, "Height (ft)": 1732, "Year Completed": 2018, "City": "Beijing", "Country": "China", "Primary Use": "Office"}, {"Building Name": "TAIPEI 101", "Height (m)": 508, "Height (ft)": 1667, "Year Completed": 2004, "City": "Taipei", "Country": "Taiwan", "Primary Use": "Office"}]
  },
  "mark": "bar",
  "encoding": {
    "y": {
      "field": "Building Name",
      "type": "nominal"
    },
    "sort": {
      "field": "Height (m)",
      "order": "descending"
    }
  },
  "x": {
    "field": "Height (m)",
    "type": "quantitative"
  },
  "color": {
    "field": "Primary Use",
    "type": "nominal",
    "scale": {
      "range": [
        "#1f77b4",
        "#ff7f0e",
        "#2ca02c",
        "#d62728",
        "#9467bd",
        "#8c564b",
        "#e377c2",
        "#7f7f7f",
        "#bcbd22",
        "#17becf"
      ]
    }
  },
  "tooltip": [
    {"field": "Building Name", "type": "nominal"},
    {"field": "Height (m)", "type": "quantitative"},
    {"field": "Year Completed", "type": "ordinal"},
    {"field": "City", "type": "nominal"},
    {"field": "Country", "type": "nominal"},
    {"field": "Primary Use", "type": "nominal"}
  ],
  "config": {
    "view": {
      "continuousHeight": 400,
      "continuousWidth": 600
    }
  }
}
```

3. Media AI

 Creative Function  Film Industry  Midjourney Model  Image Generation

3.1 Cinematic Alien Art Generation

Let's generate some alien geometries, organic objects, and landscapes using AI! We will use the Midjourney model to generate these images.

Guide

Use camera and lighting parameters, complex geometry, organic shapes, sense of motion, and style transfer to generate cinematic alien art.

Cinematic alien geometries

Here is a Midjourney AI model prompt which reads like a professional photographer and mathematician good at geometry jammed together to write.

Midjourney Prompt

```
Light Art, Cinematic, 35mm, Perspective, DOF, Field of View,  
Double-Exposure, Hyperdimensional, Icosidodecahedron, Hyperzonohedron,  
Glassy, Mirror, Ray Tracing Reflections --ar 16:9
```

The results are unique and comparable to a professionally shot cinema frame.



To appreciate the intricate detailing we rendered a variation at higher resolution. The workflow to go from an idea in our head to engineering the prompt to generating this completion requires practice and can be accelerated with tooling. We used the awesome [Promptomania](#) Midjourney Prompt Builder to pull this off without a professional photographer or a mathematician's help 😊.



Organic alien objects in super detailed 3D

Here is an example of how variation in the prompt, knowledge of how the generative AI model treats each keyword or phrase (hyper realistic, super detailed), appreciation of artistic styles (Harris Shutter), some creative experimentation of what concepts might mix well (seashell, icy, ornate, ocean, origami), helps create a alien-world organic-like object never seen before, yet familiar.

Midjourney Prompt

```
Modular Origami, 3D, CMYK, Tesla Coil, Electroluminescent Wire,  
Ray Tracing Global Illumination, Depth Map, Icy, Ocean, Seashell,  
Ray Tracing Reflections, Anaglyph, Harris Shutter, Tone Mapping,  
insanely detailed and intricate, hypermaximalist, elegant, ornate,  
hyper realistic, super detailed
```



Notice the intricate details in the completion.



Mesmerizing alien landscapes

You can also chain prompt creation by providing requirements in plain english to generate a prompt which in turn is used to generate the final model output.

ChatGPT Prompt

Act like a professional artist and midjourney advanced prompt engineering expert to create a prompt using most of midjourney's advanced

```
capabilities to generate a super intricate, detailed, highly professional art  
about a mesmerizing alien landscape. #(1)!
```

- I.💡 You can replace `a mesmerizing alien landscape` with any other topic of your choice to generate a Midjourney prompt for a different subject.

Which generates this completion. Note the detailing in English descriptive text as apparently ChatGPT is trained on a more detailed variant of prompt style compared to the above two examples.

✍ ChatGPT Completion >> Midjourney Prompt

Title: 'Ethereal Dreamscape' Generation type: Image Generation Style: Surrealism Subject: A dream-like landscape filled with strange and mysterious elements, floating in an endless void. Instructions: Use a mix of abstract shapes, fluid forms, and organic textures to create an ethereal atmosphere. Include a central focal point such as a large glowing orb or surreal structure surrounded by smaller elements that drift and swirl. Experiment with lighting effects and a range of colors to add depth and energy. Advanced: Stylistic transfer (surrealism), image manipulation (warping, blending, layering), 3D generation, and animation.

Which generates a more complex image from Midjourney when compared to other examples.



Here is a more detailed view.



3.2 Interview with Stability AI CEO Emad Mushtaq

We use ChatGPT to generate summaries of the 1.5 hour interview with Stability AI CEO Emad Mushtaq. The full transcript is available below and as you scroll down, you can see the summary of each section. We have summarized about half of the interview, and we will continue to add more summaries in the future.

Emad Mostaque and Peter H. Diamandis discuss everything from AI-generated content and property rights to ethical implications, along with the upcoming hyper-disruption wave of technology in all industries.

Emad Mostaque is the CEO and Co-Founder of Stability AI, a company funding the development of open-source music- and image-generating systems such as Dance Diffusion and Stable Diffusion.

Peter is on a mission to solve the world's greatest problems. Named by Fortune as one of the "World's 50 Greatest Leaders," Peter Diamandis is a Father, Founder, Philanthropist, Humanitarian, and best-selling author. Peter has gathered over \$100m to solve humanity's greatest problems. This show is Peter's effort to share everything he knows about changing the world.

Transcript

Transcript Summary

I am here with Imad Mustak, the founder and CEO of stability.ai. Imad is focused on amplifying humanity's potential through AI and is known for his text to image model, stable diffusion, which was released in 2022 and received a lot of attention in the developer world. Imad has a diverse background as a hedge fund manager, autism researcher, technology initiative leader, and X prize competitor. He is now pursuing moonshots that could transform many industries.

so again like if you're entrepreneurial be an entrepreneur in this if you are someone you can communicate you communicate this to other people and you get paid a million bucks a year as a consultant right you organize information if you're an artist if you're creative if you're not a tool you become the most efficient artist in the world when you lean in on this you know like systems can be out competed it's like the example of the steel mill right they're a big vertically integrated steel mills that are competed by lots of little steel mills sure micro Mills right the big corporations the big programs the big things will be out competed by just individuals and small groups building on top of this technology can do anything and a massive transformative purpose is what you're telling the world it's like this is who I am this is what I'm going to do this is the dent I'm going to make in the universe welcome everybody Welcome to moonshots and mindsets I am here with a old friend and a new friend uh Imad mustak you might welcome it's a pleasure to have you here thank you for having me we're going to have a such a fun wide-ranging conversation across everything that is of I think importance to anyone listening to any entrepreneurs any CEOs any government leaders any kids everything that you're doing is is really transforming the world let me do a proper introduction Imad mushtaq is the founder and CEO of stability.ai focused on amplifying Humanity's potential through AI you probably know stability AI because of its text to image model stable diffusion released in 2022 which rocked the developer world and I think you know broke the internet is a good way to describe it previously emad has been a hedge fund manager an Autism researcher now has led multiple technology initiatives across multilateral organizations and governments and an X prize competitor today emad is pursuing an incredible moonshot actually a series of moonshots that could transform many Industries you know

Transcript Summary

I was discussing the industries that could be disrupted by AI and the answer was that all of them could do better. I believe that this is not just hype and that the advancements in AI technology today are truly unique. AI is essentially information classification and we've come a long way from classic AI, where information was inputted and models were created, to today's AI that has the ability to pay attention to important things and understand the meaning behind data. This is called semantic understanding and it is what allows us to draw amazing conclusions quickly. I personally entered the field of AI by working on drug repurposing for my son who has autism. By reducing the amount of noise and using principle-based analysis, AI can now do what humans do by filtering information and understanding interconnectivity. This is the revolution that is happening today with personalized data sets that are good for the individual, company, and context.

I started by saying which Industries are you looking to disrupt and your answer was all of them could do better yeah and I think uh that's not hype I think that's actually from what I understand what we're going to explore true you know uh scientists and technologists have been talking about AI for decades but today is different isn't it yeah I mean we always say this time is different but this time really is to everybody over here I mean if you kind of look at it AI is basically information classification and we had classic AI which was information goes in and then you extrapolate from a data set you create really custom models and it goes out like the big AIS internet 2 was Google and Facebook taking all that big data and then targeting Peter with like ads for Rockets you know and things like that because then you know like rockets um in 2017 we had a bit of a step change where there was a paper called attention is all you need about how to teach an AI to pay attention to the important things fascinating and learn principles yeah printable based analysis yeah so it's not good old-fashioned AI or kind of logic causal based AI but it's kind of that because like um there was a book who was it by um type one type two thinking I can't I can't yeah so you know there's the very logical thing there's a freaking tiger in that bush over there yeah you know and we didn't have that second part we didn't have the ability to kind of just lead to conclusions principle-based heuristic based thinking you know this is kind of the mindset thing right whereby you construct these things it allows you to go very fast to just amazing conclusions that's part of what makes humans humans yes it's here now and it actually works leaping to conclusion but so it's a longer extrapolation based upon data it's actually learning to understand the meaning hidden data exactly it's kind of semantic understanding kind of literally it's kind of called latent spaces hidden layers of understanding so like um I'll give you an example one of the ways that I kind of entered AI was working with AI to do drug repurposing for my son who has autism yes so he was two years old at the time and the doctor said nothing could be done

Transcript Summary

I was trying to figure out information about autism studies and what was causing it. I realized that the information was hidden and couldn't be found just by scanning the data. I used AI to do natural entity recognition and built a bimolecular pathway analysis model of neurotransmitters in the brain. This helped me understand that there are two neurotransmitters in the brain and different compounds can affect them in different ways. I also learned that humans are filtering entities and our brain creates a simulation to allow us to focus. By reducing the amount of noise, AI can now filter and pay attention to information just like humans do. I believe that AI can now create personalized data sets that are good for individuals, countries, companies, and their context. I also believe that AI can revolutionize the way we find information and that humans learn by principle-based analysis and are story-based animals.

and I was like of course there's something can be done like let's try and figure out information because what we're looking for is information at each common time it's kind of the Claude Shannon theory of information theories information is valuable and as much as it changes the state so how do you find the valuable information amongst all those autism studies and figure out what's going on because everyone no one can really explain to me what caused it so just one second there because typically people are looking for a very clear-cut answer in the data that's obvious but there are answers but it's hidden it's hidden you know it's kind of hidden and so you have to look at things from a first principles basis when you can't see things just from scanning everything you have to dig down to the net nonetheless and then lay another layer and kind of when I dig down I kind of used AI to do um natural entity recognition look at all the different compounds all the different trials that were tried Builds on the team then built a bimolecular pathway analysis model of neurotransmitters in the brain so it's like just like a common cold it seems like something very similar is happening it turns out there's two track neurotransmitters in the brain one of them is Gaba you pop a Valium and chill out it chills you out and another one is goose mate and glutamate kind of makes your brain accelerate and so you know when you're tapping your leg and you can't focus and concentrate that's what happens because humans are really a filtering entity as it were we have so much information in the world it's just too much so our brain creates a simulation that's why we've got the optic nerve for example filled in instantly we are kind of constrained a little bit and that filtering allows us to focus we all know what it's like when you're tapping your leg because there's too much stuff going on so with kids with ASD they can filter for various reasons but it was different reasons not enough Gaba too much glutamate when you're born gabri and glutamate are actually both excitatory and oxytocin flips the switch on Gaba interesting so we started digging down into kind of the things and we're like ah different compounds can affect this in different ways but then what was the Upshur of that and how is this long story coming to a conclusion because there's too much noise my son couldn't speak because he couldn't formulate the hidden layers of meaning and connectivity on Concepts so kind of we've got cup here yes and he can cup your hands you have a World Cup which England hopefully will win you know and you've got um various other meanings of the word cup you form this latent space of him connective meaning to that so it can be applied in different things the principle-based analysis by reducing the amount of noise you could then filter and pay attention and build that out the AI can now do the same thing so you don't need to have huge petabytes exabytes of data anymore with structured data it can figure out interconnectivity and I want

to come back to that because that's the revolution that that you're creating right now just you know these these personalized uh data sets that are are good for you your country your company yourself your context exactly because humans are heuristic animals we learn by principle-based analysis and we're animals that are story based so you have multiple stories that make you up you know from your work on the express do humanitarian to you know bold VC fund and all the things we all kind

Transcript Summary

I am discussing the impact of stable diffusion, a product of my company, on the world of technology and entertainment. Stable diffusion is a community effort to build an AI that allows anyone to create anything from words, with the capability of creating anything from a high-resolution render to a movie. With the recent advancements in technology, we are on the verge of being able to render a virtualized world in real-time and eventually create movies by just describing what we want. This is a significant evolution in humanity's ability to communicate and create, as it eliminates the need for traditional forms of communication such as writing and clicks. The future holds the potential for a world where we can create anything we imagine, iterate it dynamically, and have new forms of communication.

of form connectivity there but those stories are very hard to map initially now we can do it dynamically and we can start building tools to re-augment human potential by doing this because we can finally have that assistance we're going to go so much have so much fun on these conversations but I want to take us back to some of people around the world for the last year or so have have heard of Dali and Dolly too and here comes stable diffusion that is an open source version of of Dolly and Dolly too and has as I said broken internet you sent me an image of the speed of growth on GitHub yeah of uh of people using stable diffusion and it was so funny because when you first sent me this image right here you see it on GitHub the the users who are uh like on ethereum over time yeah right and then there's this line that goes straight up that I thought was one of the axis for the graph yeah like what happened there so first of all what is what is stable diffusion it's one product of your company right well stable diffusion is a community effort to build an AI that allows anyone to create anything so words go in and then anything you describe comes out okay forbidden say okay uh so what we did is uh you know the collaboration with various entities uh which we paid an important part in stable diffusion to is kind of led by us we took a hundred thousand gigabytes of image label past two billion images and created a 1.6 gigabyte file that can run offline in your MacBook and create 2.6 gigabyte file which is relatively you can transmit it over over the phone network you transferred over the phone network but you don't need to transmit it once yes everyone having this tiny file that basically compresses the visual information of a snapshot of the internet can create anything from a high resolution render of a apartment to Robert De Niro's Gandalf or anything else and they can now do it in a second cent in a week they can do one of those images in 30 seconds of a second 130 of a second we've just broken real time we had a 30 time speed up in the last week so so when you mention that just as we're getting ready for this I said so you're basically able to render a virtualized world in real time yes and that's one step away from being able to Virtual or to render a movie in real time yes so we are you are we in the verge of basically and in you know we're sitting here in in Los Angeles and you've been in the Hollywood world I mean are we talking

about rendering Motion Pictures by just describing writing reading a script and having some you know stable diffusion 3.0 create a film in real time well who needs to write a script get the eye to that that's right you just say it like I want something to make me happy right and then it'll pull together various models and it can live generate a movie yeah so we're live generating so Hollywood so in the next five years yes I think that Ready Player One Oasis World minus the microtransactions of whining teenagers will be here create anything you can imagine iterate it you know dynamic new forms of kind of communication new affordances like all these clicks and things like that you don't need that anymore like this is one of the biggest Evolutions in humanity ever because the easiest way for us to communicate is what we're doing now and having a nice chat right back and forth then the Gutenberg Press came and suddenly you could communicate through written performance harder yes you know

Transcript Summary

I am discussing the impact of AI on various industries and how it is transforming the way we communicate and create. With the recent advancements in AI, we now have the ability to generate visual content, movies, and other forms of communication instantly by simply describing what we want. This technology is democratizing the creative process and removing the barriers that existed in the past. Hollywood and other centralized organizations will be disintermediated as communication and information classification becomes more dynamic. I also mention that the most successful organizations are complex hierarchical systems where people work together towards a common vision. This can be achieved by having a unifying, emotionally charged vision and good communication.

but still you see now gpt3 and other of these AIS have made it easier now for anyone to do anything like chat GPT just came out it's amazing yes everybody's making extraordinary claims about what it's going to do to Google exactly well this generative search engines is a very interesting thing um but then visual communication is the hardest you know like you toil to be an artist or all of a lot of people on this call how much have we toiled with PowerPoint yes we're gonna rid the world of the Tyranny pathway but that's my expertise from how important you know PowerPoint expert you're a communication expert so we remove the barriers whereby any visual medium can be created instantly and you just describe to it how it shifts and because it's got these latents it understands happier or sadder or you can say flamboyant and that's not a word but it figures out what that kind of means but the ability to rapidly iterate it and say okay I don't like that movie scene make it you know happier make it brighter make it more dramatic and then being able to render any form of communications like you just said you know a movie and have a thousand variants I mean what's going to happen in Hollywood I imagine Hollywood quite disintermediated yes Hollywood and everyone's thing I mean what is Hollywood right Hollywood emerged because it was far enough away from the East Coast the IP laws didn't apply right and then kind of it got constructed around here as an entity that extracted rents from performers ultimately and creatives like having creatives like the Hollywood or the music industry or others not many because they tend to be treated quite badly some people achieve Superstar them but a lot of this technology now is just truly democratizing in that things are going from centralized to the edge and we had to create centralized organizations as a people because we didn't have the information classification and communication tools to be more Dynamic it's called the

representative democracy it's neither true democracy it's representative democracy exactly at the same time you know everyone speaking at the same time doesn't make sense but what we need like the most successful organizations are complex hierarchical systems there's groups of people working together Loosely Bound for a bigger story and this is what we can achieve as Humanity like the human Colossus when we all come together to do one big thing but it gets blocked by a lot of the stuff because we're not communicating properly like one of the things I tell my team is that you know road maps are not about resources the right communication if you can communicate properly and something's good you'll always get resources for anything it's also about creating a common Vision that everybody's aiming towards in order to get there right so you're not Diversified in Thousand directions So like um Google is full of amazingly smart people right they did an amazing study called Product Aristotle yes because it's like why is one smart team better than another what was it finding the finding was it came out of two things common narrative and ideally something that kind of engages you and there's a bit of sacrifice that you know like the sort of Saleem and I and slim Ismail who you know well who sent his records talk about is a massive transformative purpose having a unifying really compelling emotionally charged Vision that you're heading towards something deep and something that you've put yourself into as well so it becomes part of your story yes as it were yeah

Transcript Summary

I talked about the importance of having a common narrative and psychological safety in a team to achieve success. I also mentioned the advancements in AI, like Galactica, that can help with creating new data and enabling more divergent and original thoughts. I also talked about monitoring blood glucose levels to maintain peak vitality and longevity. I mentioned that everyone responds to different foods differently and that monitoring blood glucose levels can help with avoiding health issues caused by prolonged hyperglycemia.

and the second part was psychological safety which I think was very extra team amongst the team the ability to actually express yourself and talk about something without fear of reproach that's what makes the most successful teams because we're too scared often about kind of our status and worrying people and things like that that's also on the research side of things primarily but I think some of these lessons kind of come because when you feel comfortable as a community and you're working towards massive transformative purpose you can do a lot more because when you're falling behind you can communicate it and you're not scared of people judging you and if you have an idea that considered you know Divergent from the from the center you feel you feel open to being able to share it in fact it may well be the right idea and I talk about the day before something is truly a breakthrough it's a crazy idea and if you're scared about actually putting forward a crazy idea then you're stuck well again it comes back to information Theory Sharon style right yep information is only valuable in as much as it changes the state yes right everyone's on the same page with everything flat equals dead exactly and this is a time when we basically have to have exponential progress and now we actually have the tools to help us to make exponential progress like you know Facebook released Galactica recently which is their language model trained on 42 million science papers they made some claims that were a bit high peak and then people like oh you can use this to create racist science papers and so they were forced to take it down but I mean it's an amazing piece of tech and we're going to help

re-release it um and this is an interesting thing you can use it to do things like a null hypothesis Creator you can use this AI to do all the sorts of things that enable more Divergent original thoughts or creativity or any of these other things like classically I couldn't because it was out of mode yeah it was out of the data set yeah of course if it's not in the data set then you're screwed where's now you don't need it to be in the data set you can create new data yeah that's extraordinary this episode is brought to you by levels one of the most important things that I do to try and maintain my Peak vitality and Longevity is to monitor my blood glucose more importantly the foods that I eat and how they Peak the glucose levels in my blood now glucose is the fuel that powers your brain it's really important High prolonged levels of glucose what's called hyperglycemia leads to everything from heart disease to alzheimer's to sexual dysfunction to diabetes and it's not good the challenge is all of us are different all of us respond to different foods in different ways like for me if I eat bananas it spikes my blood glucose if I eat grapes it doesn't if I eat bread by itself I get this prolonged spike in my blood glucose levels but if I dip that bread in olive oil it blunts

Transcript Summary

I believe that AI is a powerful tool that can help us solve some of the world's biggest problems. However, the current data sets are biased and focused on manipulation. To address the fear of AI, I believe that organizations themselves are a slow and dumb AI that feeds on us and turns us into cogs. I think this technology can help defeat that by giving people agency and allowing them to achieve their potential. The AI discourse has been focused on supercomputers, but it should also focus on ethics and the balance between red and green teaming, much like cryptography.

it and these are things that I've learned from wearing a continuous glucose monitor and using the levels app so levels is a company that helps you in analyzing what's going on in your body it's continuous monitoring 24 7. I wear it all the time really helps me to stay on top of the food I eat remain conscious of the food that I eat and to understand which foods affect me based upon my physiology and my genetics you know on this podcast I only recommend products and services that I use that I use not only for myself but my friends and my family that I think are high quality and safe and really impact a person's life so check it out levels.link slash Peter we give you two additional months of membership and it's something that I think everyone should be doing eventually this stuff is going to be in your body on your body part of our future of medicine today it's a product that I think I'm going to be using for the year ahead and hope you'll consider as well you know when you go to your website I love the notion of AI by the people for the people and then you say stability AI is Building open AI tools that will let us reach our potential let's talk about that because I you know there's a lot of individuals that you and I both know that are fear mongers around AI you know it's the devil it's going to destroy us it's going to superhuman AI is the end of humanity as we know it and I mean my position is AI is the single most important tool we're ever creating to solve the world's biggest problems and we can't solve our problems from where we were before but these are the tools that are going to allow us how do you so how do you address the Bill Gates the Elon to the World on that side the fear side well I think fears are valid because of the most powerful technology we've ever created and it comes from us but then who is us right like if you look our current data sets they're massively biased they're

fixed towards the internet and they're fixed towards manipulation the way I kind of look at AI is that organizations themselves are AI they are slow Dumb AI that feeds on us and turns us into cogs in fact I think this is concept of moloch right from the Ginsburg poem you know uh Hal talking about this carthaginian demon that pervades our organizational structures and turns us into these cogs that feeds on us effectively I think this is the first thing that can actually defeat that this particular technology that we have today because is the world happy now how many people in organizations are happy we all know you know we should talk about my idea for a happiness X prize but that's a different government so the happiness X prize but what has happened is happiness is agency happiness is achieving your potential and it's kind of going out there so we need some help but I think a lot of the AI discourse has been focused on gigantic language models and freaking supercomputers which we still have to create something that will equal and surpass us it's very religious in its way and you see parallels to religion across this sure in that you have the kind of declaiming people who like for it the people who call it heretical and then you can't even have most of Ethics in AI it's not actually ethics instead it's Ultra Orthodoxy as it were like in Islamic terms everything is haramonous as declared Halal you know classic Jude Orthodox Judaism et cetera so the same thing people look at Red teaming they don't look at green teaming right and they look at this technology being too powerful so just like cryptography we should

Transcript Summary

I am discussing the fear and skepticism around AI and its potential consequences. I believe that AI should be accessible to everyone as an open source infrastructure for humanity, rather than controlled by corporations. I think that humans at their base are good and that with more powerful technology, they will make the world a better place. However, it depends on their perspective. I mention the concept of altruistic evil and the distinction between positive and negative liberty.

keep it from the people and we definitely shouldn't give it to like emerging markets and people who aren't smart enough to do it and this would be very interesting because again if you think about the power of any believe it's powerful then the question should be what is this and I believe that this AI is infrastructure Clayton Christensen uh kind of The Departed mentor of mine had an amazing quote which is that infrastructure is the most efficient means by which society stores and distributes value obviously that's ports and things like that sure but it's also information sure and I think that's valuable when you can find it with the Shannon Theory so this AI is infrastructure for the next generation of human thought and what does that mean it should mean that it should be a Commons that is accessible to everyone it's kind of LeapFrog and you made a very definitive uh decision to make this an open source movement here yeah um your community is how large now I think we've got about 120 000 amazing in the communities and we created communities across verticals classical open source crew is product related where you have mongodb or something like that and then you go up there whereas we said language Healthcare bioml audio let's get all the people who are fantastic in the private sector public sector actually be independent together and let's jam on how do we build a Next Generation infrastructure the way I put it is let's go to the Future and bring back AI with us nice and we can choose if it's a panopticon controlled by corporations like web2 or you can choose if it's open and infrastructure for Humanity and if it's infrastructurally

marketing it's an important point because I'm just finding into that a lot of the naysayers around AGI and Asi and others I would agree with them if it's controlled by corporations which are this type of weird entity that we've created like YouTube optimized for extremism because it was engaging so ice has co-opted those algorithms and they adjusted eventually but it was slow if it comes from us and it's bias and we've broaden the conversation I think AI is more far less likely to kill us all if it ever to become sentient which is a big question you know especially if it's representative so I want to come back to this in Greater detail later but I I think we share a common belief that Humanity ultimately is good yes and it's fundamental level because that's a very important uh distinction if you believe that that humans at their base are good and you're enabling humans with more and more powerful technology they're going to be using that to make the world a better place and solve problems on the whole on the whole up but it depends on what the perspective of humanity is you know you've kind of got uh what's it called the thing when you go to space and you look at Earth yeah the overview effect the overview effects yeah you know like hopefully more people get a space and have that because people are very narrow and they view themselves like that they're Rabbi sax um former Chief Rabbi of the UK had a very wonderful concept called altruistic evil those who actually do even believe they're doing good interesting and you see that like yeah you know if you talk to especially in the religious uh religious or anything like that I'm like Isaiah Berlin's conceptualization uh British philosopher from like the 1940s he had a conceptualization of positive Liberty versus negative Liberty so negative Liberty was the freedom for anyone telling you what to do not to kind of less effect

Transcript Summary

I am an advocate for the potential of AI to solve the world's biggest problems. I believe that AI should be accessible to everyone as an infrastructure for the next generation of human thought. I founded Stable Diffusion, a company that is building open AI tools for this purpose. Our community is 120,000 strong and spans across various verticals such as language, healthcare, audio, and bioml. We have products such as GPT Neo, Dance Diffusion, Open by ML, and DNA Diffusion. I am also passionate about transforming the healthcare industry and understanding human longevity. Conventional medicine often views humans as generic, but I believe that humans are individual and should be treated as such.

capitalism and things like that positive Liberty was ability to believe in an ism something big like fascism communism capitalism islamism Etc and people use that as excuses meaning they're doing good to actually do bad but then what how does this all relate to this I believe people inherently want to do good it's just that what good is can become misdefined and co-opted and so my take was if we start building infrastructure where people can see bigger perspectives because they get the information they need what does that look like like what if we mapped all the religious decks in the world so that a child in Egypt could see Judaism for the perspective of a child in Jerusalem we have the technology to do that what if you could automatically Translate tea party tea party republicanism into libertarianism and have commonality there can we finally have the technology to do that but we didn't before so people remain in their huddles yeah they look at the other and they're encouraged to do so we've seen that increase in political polarization so Society is a typical and by the way I mean that's been what social media has effectively done so efficiently because it's beneficial for the algorithms and for the slow Dumb AIS of the

corporations that drive it let's come back to the company and the products uh just to lay out the tapestry here um stable diffusion is one product vertical area what are the other ones so for example our Luther AI Community um has GPT Neo and X it's the open source version of gpt3 by openai the most popular open language model in the world has been downloaded 25 million times incredible and so can you take it you customize it for yourself no permission needed right uh harmonize has dance diffusion which is the most advanced audio model in the world uh you'll be able to create your music so you put your own music in it and then you have your own music model that can create more music of your style and you can just basically produce it produce your own concerts extrapolate in any direction you want exactly we have open by ml where we kind of have open fold and Libra fold protein folding DNA diffusion for DNA protein matching and I've been Ella so bio LM for again some protein kind of stuff so in you know we've spent some time talking about uh a passion that I have and and you share which is the ability to truly transform the healthcare industry and in human longevity understanding why we age maybe why we don't have to I think the why is the important part right most of medicine so again when I my son was diagnosed and the scratching a wall to his fingernail spread you know eventually went to mainstream school thanks to interventions and I want to come back to those for those who are listening who have a child with autism or know someone I want to come back afterwards what were your learnings and what's your advice yeah there's a lot in the space that we kind of occupy of course um but like what I realized is that conventional medicine and a lot of things view humans as a gerdic a thousand tosses of a coin the same as a thousand coins cost at once twins coins tossed at once um but humans are individual so for example a good proportion of humans have a cytochrome p450 mutation in their liver which means they metabolize things like fentanyl yes it's far quicker yeah it's a very simple s p test but we don't do it so we just prescribe everyone the same thing and then a bunch get addicted you know because we are individualistic

Transcript Summary

I believe that AI is the most important tool we're creating to solve the world's biggest problems, but it also raises fears about who is controlling it and if it will destroy us. I believe that people inherently want to do good, but what is considered "good" can become misdefined. I think AI should be an open infrastructure accessible to everyone, so that it can help people achieve their potential and make the world a better place. In healthcare, I believe in personalized medicine and that every child at birth should be sequenced and have access to their personal AI. I see AI playing a big role in healthcare and education in the future and believe that it should be open, so that it can be used as valuable infrastructure. The ultimate goal is to help people, especially children, achieve their potential and be happy.

kind of creatures this is why I went to the first principles thinking kind of approach on this and the question is this personalized medicine thing has always been kind of out there we've not been able to reach it again I think the technology we have right now enables personalized medicine we've seen things like crispr obviously and others but more than that it's about data availability sure and viability but we do need to get to a point where every child at Birth is sequenced and that is plugged into their personal Ai and they understand exactly every food every medicine in every aspect of our living affects our physiology again you go to the Future and you bring

back the AI with you and you say what should it look like so you take a whole country and you say how do we build an amazing Healthcare System education system Etc will is there any doubt that Healthcare and education will have ai at the core in 20 years zero well I don't I think it's not 20 years it's 10 years I know but let's just say 20 years you know you and me we're like now now let's say 20 years right but then is that AI open or closed yeah and is it better for it to be open or closed yeah there's no question open is fundamentally critical open is fundamentally critical because then we build it is infrastructure it is valuable like the way that I actually Orient my rights um vinay Gupta as one of the ethereum guys think you probably know him big thinker might be crazy um he had a very great conceptualization of rights which I agree with which is the rights of children and so like effective altruism and all of that looking at people a million years in advance it's kind of difficult and it comes down to utilitarianism all sorts of weird stuff um but the rights of a child rights of a child today yes today what does that child have the right to achieving their potential yes what infrastructure do we need to give that child in a refugee camp or in Brooklyn or in Kensington to help them achieve their potential and this goes back to what's on your website in terms of using AI to help in this case a child in in the broader case Humanity achieve its potential achievements I mean that's everything yeah and so for me it's about I've oriented that on agency and happiness and it sounds fuzzy but it's literally when you have the tools to be able to do anything you know you can do anything people underestimate their agency sometimes you've got to go for your shot as it were yeah um but again it's information what information can I bring to that person what tools can I give them so that they can be creative because we lose that creative spark as we get older right yeah now it's coming back of it you know what can I give them so they can access the information they need to be educated and what's the optimal way to teach linear algebra and that's the way it's all about Play It's All About play it is all about play happiness play all the neurochemistry of your brain is maximized for learning retention experimentation around that it's float it's flow so I used to be a video game investor that was my big sector it was one of the biggest in the world and I used to judge video games by time to fun flow and frustration and so if we're building systems for Humanity we have to look at fund flow and frustration yeah because if we can get those you know when you're just learning something like wow this is amazing and how do you get in there it absorbs amazingly quickly but our education system is not set up for that our Healthcare System

Transcript Summary

I believe that AI should be accessible to everyone as an open source infrastructure for the next generation of human thought. This way, it becomes a Commons that everyone can access. I think AI is more likely to be beneficial to humanity if it is open and not controlled by corporations. I have many products in different vertical areas such as language, healthcare, audio, and others. I am passionate about using AI to transform the healthcare industry and help people achieve their potential. I believe that in 20 years, AI will be at the core of healthcare and education. I am working towards building systems that maximize happiness, creativity, and learning. I am also exploring the issue of intellectual property rights in AI and believe that these models will lead to significant IP on the invention side. I am constantly working to improve my products and bring the latest technology to the market.

definitely isn't and you've got a messed up Healthcare System yeah it's maximizing for frustration yeah that's where I speak about it openly saying my mission is to crumble and destroy the healthcare system and also education system which is uh which is really sad so again it's very interesting you look at the US and inflation numbers education and Healthcare massive inflation yes everything else not really yeah it's a percentage of the other person's income is it's an and it should be going to zero right the top health care and the top education should be all II driven it's all basically the cost of electrons life expectancy is falling in the US a little bit over the last couple years before suicide before covert it was falling yeah I mean this is the thing whereas again it's not complicated but it does require coordination so the question is can you create shelling points so for me open source software this next Generation this model-based one whereby stable diffusion is basically a programming primitive just like you have a library to do various things for the programmers out there so in the good old days when we started programming we used to have to code everything by hand now you've got GitHub and things like that it's more like playing with Lego right yes and you're sticking it all together it's the new type of thing a 1.6 gigabyte file that is hashed so it can be common across every single computer in the world that you can call something in an image comes out and you know predictably what that is no matter where you are and you can also take an image and put text on the other way yeah that changes the Paradigm but then what if you have that for language audio all these different modalities okay I'm gonna go I'm gonna go here next then on that which is what happens to literal property rights IP rights who owns the IP of those images is it the person who puts the prompt in is it you know help me understand where that where that evolves to that'd be nice nobody knows okay it's a fair answer yeah I mean nobody knows I personally think it should believe it should belong to the person that prompts it because again can we put this out as a Commons for Humanity yeah um like we did put an ethical use license around it for various reasons that will be replaced by a purely open source license but again I'm doing this as building blocks right it's really the person that has the action because these models do not have agency if you're gonna breach copyright you can type in I want a picture of Mickey Mouse and then if you sell that done but it's like photos it's like telephone it's like Internet it's like Photoshop this is a tool but it's a tool of a very different type because like something like Photoshop is a million lines of code this is a binary file that can do anything that photoshop does but you still need to act on it last year there was a patent awarded to an AI in South Africa which was interesting I think more of a gimmick than anything else uh do you imagine these uh these models will lead towards significant uh intellectual property on the invention side of the equation oh hundred percent they already are if you look at Google's new TPU chips they're partially built by AI that they built sure yes I know like we just released um open uh LM from our copper lab which is an evolutionary algorithm for code for robots so we're trying to optimize robotics via that and then when we bring in video we'll have even better robots

Transcript Summary

The speaker is discussing the impact of technology and how it is changing the world. They believe that with the advancements in AI and open source software, there will be significant disruptions and growth in various industries. The speaker mentions how the creative industry and video games will change, and how the healthcare and education systems need to be re-evaluated. The speaker also talks about the importance of understanding the changing world and the shift in value and how it will affect individuals and companies. They mention how the models of AI will allow for seamless creation of new things and how the value in these industries will change. The speaker believes in the power of open source technology to be made available to everyone.

so are we going to get to a place where I'm saying listen I'm please invent a device that does this for me that's under this price point that's made a commonly available materials and over you know constrain it left right and center and then design it now go and print it for me and deliver it for me the next day we're going from mind to materialization in one sense yes yeah I mean you know again it's kind of the was it the Star Trek thing yeah it's it's a little bit longer yeah oh my god um Earl Gray tea hot exactly but then you can kind of see this already because like there's an app on the app store right now you put all your Lego out and it scans all the Lego all right nice and what can you build with it yes it comes up with this is how you build it as well so this is just an extension of that again they use a Transformer based architecture for that well I'm interested when uh it'd be fun to say okay create new life forms I mean we're not far from that well you know I'm gonna leave that to other bioethicists I don't want to stuff my foot in that one there's ethics involved in that exactly I mean the life form thing is very interesting though right because again everything is happening all at once and so it's not just biology or physics or sociology all these models all these Technologies all seem to be converging at once so you can create anything and all the barriers are dropping at once yes so and that's complicated yeah so the world is I don't do you think anybody truly understands how fast the world is about to change no I mean like look at creative industry video games 180 billion dollars a year like Disney spends 10 billion a year Amazon 16 billion year on content all of that's going to change in the next couple of years alone just from one tiny little two gigabyte file extraordinary as well as Healthcare as well as education as well as you said every single industry we're gonna so here's another question go ahead please well I was going to say industry is about again information Theory yeah that's most Industries are based on especially service-based Industries and so once you can basically take a system you can have human input in the loop to train a system that's a generalist to understand principles just drop just about everything yeah I'm gonna pause on that because it's it's probably one of the most important things any entrepreneur any CEO any parent any kid anyone needs to understand we're about to enter a period of of hyper uh disruption and growth and Hyper opportunity creation yes like what happens is that in any area you create there are value spikes so you can look at it as like a flat area and then companies and individuals occupy certain areas they've got like a mix of skills right and that's how you earn your living and these are like spikes that's all going to get shaken up and it's going to be the value is elsewhere yeah we don't know where the value will be we've got some guesses so like in a time when everyone can make anything what becomes valuable something yes so if a model that can make anything that means Disney should have their own models right to create Mickey Mouse's and things like that but now they can use that not

only internally to save money on creation they can use externally why can't we have Mickey Mouse having coffee with Master Chief at a Starbucks and microtransack pay that it's been the promise of the nfts and things like that now with these models you can do that seamlessly yeah I'm on demand yeah over and over again so one of the realizations for me is that uh this open source technology made available to everyone

Transcript Summary

I am focused on applying first principles thinking to information flow, social theory, and systems. I believe that by making open source technology available to everyone, we can create a world in which everyone has access to the best education, healthcare, and entertainment. The cost of everything will come down to raw materials, IP, and electricity. I am working on creating a stable diffusion of AI models that can be used to augment and replace existing systems. With these models, we can see a virtuous cycle of kids and AI helping each other. I am amazed by the real-time rendering of the world that is made possible by these models. I believe that we are living in a simulation and am interested in exploring the limits of this world that I am creating through the disruption of industries.

effectively has the potential to make everyone the equivalent of millionaires billionaires trillionaires when whatever you want can be manifested right you can have the world's best education for your child independent of where you live and what wealth you have you can have the best health care available you can have customized entertainment you can know it's we end up in a world in which the cost of anything is raw materials in IP which is going to be disrupted itself and electricity yeah pretty much again like you look what Elon Musk did with the SpaceX right yeah he's like let's break it down towards the constituent cost of this yeah and basically with the first principle thinking on Rocket so that's principal saying it so what I've been doing is first principles thinking on information flow social theory and our systems it all comes down to just information being in the right place at the right time to make the Maximum Impact and if we can give that as an open architecture to the world then we can augment and then replace our existing systems the better ones because they out compete you go to an African nation and you teach every child with an AI that teaches them and learns from them within a few years they will be out competing children in the top schools in New York what if you give them the ability to code the system as well and improve it it becomes very interesting the kids are helping train the AI and the AIS helping train the kids yes but then also the kids can improve the actual code that they have there yeah like this is something we've seen in a virtuous cycle a virtuous cycle and you make that open and then you make it transplantable because what you have then is you have models that are standard like a base so I like to call stable diffusion one was a precocious kindergartner then it was a precocious High School stable diffusion two stable to Fusion 3 looks freaking amazing it's going to be like a university level student so stable Fusion 3 by the way is the real time rendering no that's labeled effusion two that's two okay so again like to give you an example of the thing when we launch table to Fusion one on a top end a100 which is like a super super computer chip right and we have a lot of those we can talk about the super computers in a bit it took 5.6 seconds on August the 23rd when we launched it to render a single image run in a single image in 512 by 512 okay today it takes 0.9 seconds in 768x768 we've just sped it up 30 times amazing 30 times and I'm still I'm still Blown Away by the real time rendering of the world once you get

below 200 to 220 milliseconds of response time it opens up entirely new UI ux and we didn't just release stable diffusion as an image Creator we also released an impacting model so you could take you know Emma's hat and you could turn it red just by describing it we release a depth to image model so you can do Transformations released an upscaler so you can have 64 by 64 image and the AI fills in all the details to take it to 1024x1024 transforms the storage industry for media amazing and this is real time now as well it's and so listen I have to ask you the question but come back to it later which is do you believe we're living in a simulation uh um yes or no yes yeah as do I I think we're living in an nth generation simulation but that's a that's a different story okay we'll come back to that for folks are interested but okay how far out are you able to imagine this world that uh that you're creating the disruption of Industries and

Transcript Summary

The speaker believes that the next generation of open source software has the potential to disrupt multiple industries and make everyone the equivalent of wealthy individuals. The technology is advancing at a rapid pace and the speaker advises a 20-year-old entrepreneur to focus entirely on this area as it is expected to bring in trillions of dollars in investment. The speaker views this technology as transformative and equivalent to the impact of the Gutenberg Press, but on a much larger scale, as it encompasses not just writing, but images, 3D, creativity, and more.

transformations how many years out I don't know anymore yeah um like when I started so I founded the entire open source Art Space from January of last year when it started because he had a generator model and then open AI released this clip model which took images to text we bounced things back and forth for each other we're like wow that's the way to do things creators and discriminators as it works and it's advanced advanced sounds like this is the next big thing humans can now communicate visually right it's the next biggest thing since Google make press and it kind of went as expected plus minus six months I thought we'd get to stable diffusion in Q1 Q2 of next year and that was a big massive I built a gigantic super computer and everything to get to real time I think it would take another year or two and instead it took like four weeks and like again this is what you mentioned earlier the number of GitHub Starts Now for stable diffusion is more than Ethereum Bitcoin and just about everything else and that took them 10 years in three months yeah I mean again it's the notion that people have no idea how fast the world is changing and is accelerating and it's the it's what you came back to with the common Mission and the energy yeah when you sent your first Bitcoin it was an amazing experience they got overtaken by raccoons and stupidity yeah and they try to create alternative system outside the existing system the interfaces were all the robbery and all the profits were made this is something different whereby you're talking when we talk to developers and the contributors that are increasing in the ecosystem they're so energized and this is what drives things forward like when you see teams that do the biggest things they have the energy it's almost palpable right where it's like that driven thing but we've got people from all over the world like one of our latest developers was an amnest and warehouse worker at the start of the CEO who taught himself to code and now he's building the most advanced models in the world we've got 16 year olds to 62 year olds and it's a team of how big so our team is 137 but the but the developers well but they're teams of one right their team and their individuals are able to

use this to create oh yeah so like if you want to create with this you can just do it by yourself so um there's a fantastic Twitter you can look at levels.io and he's like I'm going to create businesses by myself that making millions of dollars of things just because you could took this primitive and you wrap things around it and you're just making money like Avatar me you can put your own face into the model 10 images you can create a Peter diamandis model and we can put you in space in fact we'll do that a bit later so listen you're you're listening your 20 year old entrepreneur you're in college you're you know finished you're skipped College whatever it is what are you what's your advice to that 20 year old listening right now you should drop everything they should focus entirely on this this is the biggest shift ever self-driving cars 100 billion dollars went into crypto hundreds of billions dollars equivalent to 100 billion a trillion dollars gonna go into this sector I don't say that again how much 100 billion the next few years and then a trillion dollars will go into this sector because it's so transformative and so few people understand it there's never been something a technological advance that will diffuse as fast as this so this is electricity this is the Gutenberg Press times a billion this is a good this is a Gutenberg Press times a billion because it's not just writing it's image 3D it's everything it's creation protein folding it create it's creativity and uh extrapolation uh so what does it look

Transcript Summary

I understand the importance of focusing on new technologies and innovations, especially in the field of AI. The rate of adoption of AI is exponential, with 80% of AI researchers working in this field in recent years. This exponential growth means that companies need to be proactive in understanding how this technology can potentially change their entire business model. Companies should have a team dedicated to exploring how AI can be integrated into their processes and procedures. The classical innovation process inside a company may be limited, so companies need to think from first principles and determine if their current format will still exist in the future. The growth of AI is exciting because it has the potential to transform the inefficiency of the world and empower individuals to create value.

like to focus on this so you're again you say drop everything in focus and I get that and and uh you know part of me is like maybe I should do that too but what does it look like to focus what would a person do so again like if you're entrepreneur be an entrepreneur in this if you are someone who can communicate you communicate this to other people and you get paid a million bucks a year as a consultant right you organize information if you're an artist if you're creative you become the most efficient artist in the world when you lean in on this you know like systems can be out competed it's like the example of the steel mill right they're a big vertically integrated steel mills that are competed by lots of little steel mills sure micro Mills right the big corporations the big programs the big things will be out competed by just individuals and small groups building on top of this technology can do anything but we and we've seen that over and over again in every ex every converging exponential field we've seen uh entrepreneurs disrupting and the rate of you know I think of this as the asteroid impact that the slow lumbering dinosaurs all die and the Furry mammals the ones that rapidly evolve yeah I mean like let's take a practical example and this is what we said earlier about chat jbt why does anyone need to use Google image search in a year yeah where you can create any image just by describing it and then iterate it just with your words

yes just attribution that's it but that's an easy lookup table yeah all right true absolutely I so now let's take it okay the 20 year old drop it start experimenting start playing start using uh now you're running a company you're running uh you know 100 million dollar company worse off a billion or 10 billion dollar company and you see this how fearful should you be and what should you be doing again you should be leaning in to understand this the classical Innovation process inside a company is very limited you should be having kind of uh crack team of people who've just given Freedom say how can this potentially change my entire company you know um but it's hard because you are fighting against the inertia of your company you're following people being used to certain ways of interacting or certain ways of Distributing stuff yeah so this is why I think you have to think and think again from first principles if this technology is real time for us to cross modalities and it can look and understand stuff let's go forward five ten years and work back does my company still exist in this current format can I out-compete that it's not how can I fold these into my current processes and procedures Etc because this is the other thing as a true exponential like the AI research actually in this area 80 of all AI researchers become in this area in the last few years and it's exponential with a 23 month rate of doubling a true exponential the adoption of this now is also exponential because everyone from around the world is using it and then each of those introduces another two people yeah it goes four and four and four so right now it's like a wave that's under the surface it just hasn't come yet next year is when it cracks and the year after is when it crashes all right I hear you and um my reaction is I'm excited as hell about that right it's like because it's about transforming the inefficiency of the world today it's about taking individuals and empowering them to do to giving them agency like you said but this is very interesting inefficiency is where value is created

Transcript Summary

I understand the concept of capturing value through inefficiencies in the current paradigm. With the advancement of technology, certain inefficiencies may disappear and the value capture will also change. This is why it's important to think about where the value will exist in the reconfigured landscape. For example, in the legal industry, the use of technology such as do not pay.com has reduced the need for lawyers in certain cases. However, there will still be a need for litigators. The movie creation and video game industries will also be transformed by technology. In the hedge fund industry, the day of the trader investing on their own without the use of technology is long gone. The advantage in trading will come down to information theory. The tools that people use for investment will change and people will be able to visualize stories better. The investment will also shift towards themes such as ESG and indexes.

in the current paradigm in some ways or what values because efficiencies exist what value is captured by inefficiencies so we have to have we go to The Gatekeepers and we give them like you pay your lawyers loads of money right you've got your accountant some other things like that because there are inefficiencies in the system so you pay them to remove the inefficiencies as it were some of these inefficiencies will no longer remain in the system but that means that that value capture will also disappear and again this is why you know you have to start thinking how where does the value look like in the new reconfigured landscape if do you have an example in an industry that exists today and just to make this concrete for someone sure you don't need lawyers anymore uh for

everything awesome I mean like you know you've got do not pay.com those are massive things they automatically write your tickets for you and kind of yes for those who don't know it's if you got a parking ticket and you go to do not pay they will figure out the legal loopholes and arguments to get you off your ticket yeah um but like you still need litigators right at what point can you have a robot litigator it's like Phoenix Wright on steroids right um you know if you've got the Again movie creation thing the movie creation is just going to transform completely yeah video games you'll be able to have your own user generated content you know like again people can right now use stable diffusion create any image then go to chat GPT it's an amazing system by openai and they can chat within like oh okay that disrupts a lot of Industries as well because it knows all the different things but it's just a model it's a blob it doesn't look on the internet what happens when you hook up these models that are principle based to the internet stable to Fusion Plus Google image search is actually incredibly powerful but people are just looking at that stable diffusion stable diffusion as part of a pro architecture is incredibly powerful people are making movies so like if you look at our friends at the corridor crew for example um they did a movie called Spider-Man everyone's home where they created a custom model based on the spider-verse movie with Miles Morales yeah a couple of days they created a three minute trailer that blows away anything big studios can do just a few people a few thousand bucks let's go to a few general questions on on AI first of all uh you come out of the hedge fund industry I have to imagine that the day of the trader investing on their own without the use of any of these Technologies is long long gone does anybody haven't you know have any advantage on their own I mean I think that's the thing what's an edge in trading right and again it comes down to information Theory what can move the Apple stock 50 percent very little information what can we have at five percent a decent amount of information one percent quite a lot of information right so we're just looking at the narrative and the incremental Narrative of this because again as humans we're heuristic creatures so like I used to be a specialist in a number of markets all Market was one an oil barrel is fungible going around the world because you can shift it on a ship sure but the impact of a Libyan Barrel going offline was the third is impactful as an Oklahoma Barrel why because of most of the money is in the west as opposed to near Libya so they feel it more and the market reacts more the market is a counting mechanism and a voting mechanism I think the tools that people will use for investment are going to change now because you'll be able to actually visualize stories better and people will kind of introduce that this is kind of why people like invest on themes like ESG and you have these index

Transcript Summary

I understand the potential disruption of various industries by AI technology. The stock market is expected to change and adjust as a result, but it's unclear which industries will be disrupted faster. The technology will certainly have an impact on the hedge fund industry and could be used to assist hedge fund managers in investing. The key to investing is understanding the story behind it and being able to de-risk the path to a terminal value. The concept of a singularity in the AI array is still up for debate, but I believe in the idea of an intelligent internet where every person, country, culture, and company has their own models that interact with each other in an optimal way for humanity. I am pushing for open-source models and have created an organization to standardize the architecture across different verticals. The idea of these personalized models is important to understand as they could lead to an emergent global consciousness.

trackers and all these other things so I think that's going to be very interesting but then which Industries will be disrupted faster than others we don't know who will be at the edge in the Forefront of this embracing this technology and then who can be left behind we don't know so I think the entire stock market's going to change and adjust it could be a period of supernormal profits and then out competition as well this is against a backdrop of inflation and recession and all sorts of crazy stuff all at the same time so will you be using this technology to assist a hedge Point manager out there on investing yeah I can almost my own hedge fund no look I mean again I think this technology would be persuasive because people again will see the power of this and ultimately like I said investing usually comes down to stories you know like for all influencing influencing human desires and intention exactly like you know if you're doing VC to fund management whatever you just basically you've got a story you can say for all you want that you're trying to be like quantitative and this and that but nobody knows the future so you construct a story but you know you're saying what is the evolution of that story going to be such that someone will buy this from you at a higher price yeah that's all that matters and part of it like you know for entrepreneurs one way that I suggest to look at things is what's the terminal value you get someone to agree to that all you're doing when you're raising money is you're de-risking your path for that terminal value and people don't realize that's all of your assessment yeah it's a very simplified view of it so I'm excited to have you at uh speaking at abundance 360 this year thank you for uh joining us in March thank you for having me um and you know I've dedicated an entire day we've added a fourth day of the program focused on AI because it's just I think people need to understand how powerful and disruptive and a view of what's coming another one of our uh our rock star speakers there someone who you know Rick Roswell and Ray's going to be joining us to speak as well uh and you know it's been raised long-held belief that we're going to reach human level a of like 2029. uh at however you want to Define that um what's your what's your thought about that do you think that that is the case uh how do you think about it I'm not sure I think it's singularity in the array um look again how have you defined I think my ideal view is that a future that's an intelligent internet every person country culture and company has their own models and they're all interacting with each other in an optimal way for Humanity so again I I kind of go to this thing of you know Tim Urban and wait but why when he wrote the article about neural link had this option the discussion of the human Colossus I would like AGI to be something of all of us working for us to make us better and I think part of this is why I'm basically pushing open source and these open source models and I've created an organization that goes and has all these verticals that will then be spin off into independent organizations so we can standardize the architecture across that and it can be an emergent Global Consciousness as it were I think it's important for us to talk about the idea of these uh these models and these these personalized models to understand what that means so when you build a massive model that is trained on everything out there it's not really necessarily useful but you talk about creating models for

Transcript Summary

In summary, the speaker believes that AI is going to disrupt many industries and transform the inefficiencies in the world. The value capture in the current system will change and the stock market will adjust. The speaker has dedicated a day at the Abundance 360 conference to AI because it is so powerful and disruptive. The speaker believes that the future of AI is an intelligent internet where every person, country, culture, and company will have their own models that interact with each other optimally for humanity. The speaker is pushing for open source models and has created an organization to standardize the architecture. The speaker has built a 4100 cluster with Amazon that has more compute than most countries and is available to everyone. The speaker is compressing knowledge by choosing the most important neurons and creating foundation models that can be refined for personal use. The speaker is training GPT4 and looking at how people use the models to compress them down even more.

Nations for cultures for companies for individuals can you just give us the 101 on that so people understand the power of that and the value of that knowledge evolves at different Paces right like there's knowledge on how to use a toilet that's been with us for many many years you know versus knowledge on Foundation models which is just very new it kind of considerate Fashions is like Pace layering as it were I think the way that models will be built is that you should look at them as like pizza bases so we're trying to figure out what an optimal Pizza base is for a generalized image model then you've got maybe an Indian model with a bit of culture in there and then Indian fashion in there and then Indian fashion for emad to try and be fashionable when he's in India and so you can kind of look at those bases because what we do is again we flip the Paradigm so to try and stable diffusion like we built a 4100 cluster with our buddies at Amazon to put that in context the fastest super computer 1000 a100 cluster is huge yeah to put that in context the fastest supercomputer in the UK Cambridge one is 640. yeah we've got about 10 times the compute of NASA well roughly speaking it's one of the top 10 in the world and next year I'll go 10 times bigger yours will go 10 times bigger yes okay and we make that available to everyone and it's on it's on the Amazon Cloud it's on the Amazon Cloud yeah because it would have been a bit too much for us to build it by ourselves um so we got them to lean in on that and kind of build this uh you know Facebook has 16 000 for example because they're pushing Big metaverse that's the fastest in the world if you want to have an idea but yeah this is an exponential thing whereby we have more than just about every single country and we use that to take 100 000 gigabytes of images and compress it into 100 terabyte yeah 100 terabytes yeah I can press it down so into two effectively so fifty thousand to one compression of knowledge all right and so when you say that compression when you're talking about a neuron that you're you're not all of the connections or the pathways neural net are useful or valid yeah right and so what you're doing is you're choosing which ones are yeah because it pays attention to what the most important and then it creates these latent spaces that you poke with the prompts which are the words that you put in right so like uh open AI for example made gpt3 it was 175 billion parameter model the next step after that deep learning because that's called Deep learning this thing yes like because we spend all the energy and all that compute you don't have to you can then take that base that pizza base you can inject some Peter then you have a Peter model it just takes 10 images and you can put yourself in anything or you can spend a hundred thousand images and do an even more refined model but it forms a basis what's called a

foundation model in a way but you can even make it even more efficient so gpt3 is 175 billion parameters so it's like 80 times bigger and they said they said gpt4 was going to be like 100 trillion parameters it could be yeah you know I think the chat GPT is actually training gpt4 right now because what happens is that you have this Foundation but then in the classical day age of Big Data What mattered was who you were so they took your data they built these big models and they targeted you well Metals now in the day of I say big models is how you use the models because not all those neurons are needed we don't need all two billion images right right so gpt3 was 175 billion and they saw how people used it and they identified the neurons that lit up yeah and they compress it down to 1.3 billion parameters so now when you're using chat GPT you're training one of those models we're looking at how people are using it to compress it down so say what Fusion you can compress down even more but then this is personalized model that's very

interesting because you can have almost this standardized Bass and then you can inject your own context into it and the context of your culture your company your community and others and so that base is manual bill in that you can extend out the latent spaces so like in stable diffusion one we didn't really filter it so Mona Lisa is over fit there's too many pictures of moon Lisa so it's very hard to get her out of the picture frame now we adjusted it so you can make her swim very easily because it's not over fit anymore and it'll continue improving and adapting so we'll have better and better bases and in a year's time we'll have a very mature base that people can take an extent so like in Japan they took stable diffusion and they adjusted the text encoder for Japanese culture and then it meant that if you use normal stable diffusion because it's very Western oriented salary man means a happy man yes in Japan diffusion is very sad man yeah very sad man indeed because I understand that context fascinating so in the future if I wanted to you know I'd coach a lot of entrepreneurs and a lot of CEO through abundance 360 and so forth if I wanted to create a virtualized version of myself that in certain circumstances would react in a certain way that's pretty easy it's coming so pocket coach across I wanted to like have Tony Robbins in my pocket on in the right moment or the Dalai Lama it's all possible yeah and that will be probably a four gigabyte file and how far is that from now let's go watch it now you could do it now and you can do it fully realistic like um you know I think I think the term is holy that's amazing yeah um because you have the volumetality you can make it indistructible for me human like stable diffusion 2 is pretty much photorealistic now stable diffusion 3 will break that barrier and obviously we can animate now so again that's kind of crazy that's crazy you know metahumans epic uh kind of epic game style or what Nvidia are doing you can also do human realistic voices with fully emotional range as well so like my sister-in-law Rana runs a company called semantic um she reconstructed Val Kilmer's voice for Val and Top Gun nice I was doing all the video games and if you go to sonantic.io you hear an AI tell you that it loves you and it's really creepy uh she assaulted Spotify so I'm sure we'll have some really engaging podcasts and things like that um and like again across all modalities now in narrow you're achieving human levels and going to human levels of performance and benchmarks from media to understanding to Output and the barriers to putting yourself in like I said you can train a model now in like less than an hour with 10 100 images of yourself to put yourself in anything for a buck because we've done the heavy lifting of millions of bucks of pre-training the model as it were yeah whereas classically AI wasn't like that so if I wanted to create a a virtualized host of myself being able to on screen play me say what I want to say that's here now that's it now like the technology is here now the implementation needs to be there so like um I mean one of the companies we work with is called

alessia.ai and so they use our language models and so you can upload your scripts and it'll learn how you speak but then the voice technology they haven't integrated the voice technology we have our next Generation audio technology that'll come in a month or two and then it will learn some of you speak literally I was using last year at a minus 360 I was using a company called Soul machines and they had a virtual humans yes all right and so I would love to create a virtual version of myself for anybody I want and that's here let's see again solar machines is upgrading now thanks to the technology that we're open sourcing this is the other interesting part open source will always like Cloud source because close Source can always take open and add in data yes and they can have very focused teams focusing on certain use cases so we're literally upgrading the foundation of all of these companies as we release these models interesting and then you mix and match and that's where the value is actually who was it who said I think it was my friend Jason or someone else no it was um one of the other VCS most of the money in the world is made by aggregation and disaggregation it just depends on which part of the cycle you're at interesting I I can see that yeah I was having dinner with Reid Hoffman I don't know a month ago or so and he said something which is interesting he said every profession is going to have an AI copilot very soon and I've been saying this for medicine I think it's gonna be malpractice to diagnose without having AI in the loop and we'll see what the time frame is five years is my guess um but I can see an AI co-pilot as an architect as a lawyer as a chef as everything right so how far is that well I mean it's here for code right now yeah right sure so co-pilot literally was called from Microsoft GitHub and then code whisperer now from Amazon they help you write better code it's about 50 speed increase amazing that's what we've kind of measured so far which is insane like you type in I want to have a piece of code that does this and boom is there I maybe need to edit it but it doesn't matter it makes your life easier right yeah uh like say what a fusion is a co-pilot for art so artists use it to iterate rapidly on different concepts and they take it and there's a Photoshop integration and they use it as part of their Photoshop process I used to say that the crowd was the interim step until AI right so GitHub was the crowd and now you've got you know yeah and you know there's questions around you know what was it trained on because it was trained on an entire snapshot of things but again it's like a different type of information flow to the web 2 economy we're going to skip over web three because it was a bit crap and we're going to go to web4 or whatever it is now and you know maybe you'll create web before I it looks like AI if you do a cool logo for that I should get that unstable diffusion to make it I've it's in fact it's listening and it's made it um so uh I love Jarvis from Iron Man yeah uh and I find uh Siri and Alexa and Google Now kind of disappointing uh how far are we from Jarvis so I think Jarvis is probably two to three years um because like right now you've got a Macbook and yeah it's the latest one M2 in front of you yep 16.8 percent of that chipset is a neural engine that's optimized for these Transformer based architectures uh stable diffusion is one of the first models to actually go down to that level and so when will we see that on this machine I think apple is basically aiming for like 70 80 of everyone to have this and you can go from Siri one to three five nice and this is why Apple's been talking about privacy and things like that because there's a new paradigm of the internet whereby the classical web 2 internet was Intelligence at the middle coordinating us and feeding on our dreams and hopes and emotions to sell Us ads now it's intelligence at the edge where by you've got your own Apple ID you've got your own privacy layer and then you've got chips that can run AI at the edge that really understand you that's why the Noggins experience is seamless and Google also realized that so they're building stuff into the pixel phones yeah I think people need to realize that the power of these future systems will call it Jarvis for lack better term is when you give it open access to everything in your life right

you let it watch what you're eating let it read it to your emails listen to your conversations because it makes the world the term I use is automatical uh in that regard yeah and you need to have the foundation models to do that because it needs generalized knowledge and then specific knowledge and contextual knowledge that can adapt to your needs this human in the loop process is very important and so there'll be big AIS in the cloud but then a lot of AIS on the edge and they'll interact with and talk to each other and brought up with each other because what these models also do is they take structured data and turn it into unstructured Data so again stable diffusion a few words Robert De Niro's Gandalf you get a photorealistic picture of Robert De Niro's Gandalf yep that's structured unstructured data yeah and then you can go both ways a brief note from our sponsors let's talk about sleep sleep has become one of my number one longevity priorities in life you're getting eight deep uninterrupted hours of sleep is one of the most important things you can do to increase your vitality and energy and increase the health span that you have here on Earth you know when I was in medical school years ago I used to pride myself on how little sleep I could get you know it should be five five and a half hours today I pride myself on how much sleep I can get and I shoot for eight hours every single night now usually I'm great at going to sleep if I'm exhausted you know I've worked a hard day I'm right out but if I'm having difficulty and it occurs I'm having insomnia my mind's overactive and I need help to get that eight hours I turned to a supplement product by life force called Peak rest now Peak rest has been formulated with an extraordinary scientific depth and background includes everything from long lasting melatonin to magnesium to L glycine to Rosemary extract just to name a few this product is about creating a sense of rest and really giving you the depth and length of sleep that you need for Recovery it's a product I hope you'll try AI it works for me and I'm sure it will work for you if you're interested go to my lifeforce.com backslash Peter to get a discount from a life force on this product but you'll also see a whole set of other longevity and vitality related supplements that I use we'll talk about them some other time but in terms of sleep Peak rest is my go-to supplement hope you'll enjoy it go to mylifehorse.com backslash Peter for your discount so let's talk about uh something which I have an opinion about I'm curious about yours I think I know it which is the idea of privacy which fundamentally people all want privacy I don't believe it really exists I mean when AI can read your lips when there's data flowing everywhere where encryption what are your thoughts about privacy and how do we deal with it you have any ideas I think so privacy you should always look at kind of what the downside to not having privacy is right actually people are more than willing to go up their data too willing in my opinion yeah everybody's sign everybody clicks that yes I accept this 15 000 pages of of uh of legal exactly and then you know like you have to think as well different paradigms like you know in China will there ever be privacy probably not and you have the social credit score and it's an Opticon being built by AI Etc and the Western parasite what's the downside on the Privacy thing like what if your stuff isn't private it's basically Bad actors using it in certain ways which can include AI algorithms trying to manipulate you I think again what Apple's doing is building a paradigm for actual privacy because it's aligned with their business model even other companies now Facebook and Google have enough information they don't need your data anymore like who actually wants your data I think is a question like we review ourselves as these snowflake like you know wonderful things who actually wants your data at this point but the systems have adopted to do that and policies adopted to that as well with gdpr and all these other things some of them overreach I think a bit but we are moving to this area whereby nobody needs your data anymore and also the sisters are now available to give you that privacy that you want and I think people want to opt in rather than opt out of a lot of different things to get more resources and other stuff finally the final

element is that Federated learning has matured now what does that mean so Federated learning is when you take the model to the data so you used to have to ingest all this data and train the models now if it's just like a freaking gigabyte you can send the model to the data it can train and then without saying it's Peter or Ahmad it can Upstream the output so we're seeing that in HDR UK for example Health Data we're seeing it with Melody which is a thing with a lot of pharmaceutical companies coming through to open source analysis of patient data you can finally get that where you don't have to sacrifice privacy to build Ai and models and that's going to be pretty amazing to again Advance the field because you have access to so much more data to build better models amazing let's talk about the the perceived downside and you know I have to imagine that um as much incredible compliments and and the world should thank you for the work that you're doing because of the impact it's going to have you're going to have to have detractors who are worried about technological employment or malicious use of AI or fake news and all of that what what concerns you and I know your your principled man who thinks about this deeply what concerns you most I don't have all the answers like and that's and that's a fairer statement to me yeah I mean like genuinely like what I saw was that very few individuals had control of this most powerful technology and then you know like it's very weird things like people like open source AI is like nukes and like so why should you control the nukes you know it was a very strange kind of thing they're like no only like it shouldn't be open source I was like so why should big companies control it like again we live in largely a democracy we live in a society and so my take was like let's educate people get this technology out there and let's have a common conversation about it because I have my own viewpoints and they're there but again I'm not representative of anyone I'm just me running my own company trying to catalyze this because I thought it was important given the fundamental change of society that will be caused by this technology now because exponentials are a hell of a thing for it to get out there and so you need to make splash so you know I've got hate mail and kind of all sorts of things because it is disruptive and we have to be aware of that it is crazy and it will cause fear and we have to be aware of that and we have to decide to get together how to do that so like for example there are artists in the data set because it's a snapshot right sure it's less than 0.5 percent and so is it ethical legal and moral to have them in there so people can prompt an art style and mash them together I think it is but does that mean that we disregard artists who want to opt out of the data set no yeah because they're part of the global community so we've built it optown and opt out mechanisms and by the way those artists are influencing other artists normally in the course of just them going to museum yeah exactly and you know what we have is now we've had like four or five thousand artists sign up for spawning half of them have opted out half of them have opted in because they'd love to see their work influence the world but how many people really absorb the param to the discussion very few so like I said my thing is that again this is fundamental infrastructure access to this technology is a fundamental human right yeah because otherwise what you're going to have this is a discussion that you know you've had many times superhumans and normal humans yeah the ability to communicate and create makes you a superhuman because just not only images like it's presentations it's being able to like we have voice to voice technology that can allow you to speak more confidently it's interesting I mean people need to realize that today the poorest Among Us in society have more than the kings and queens had you know a couple centuries ago and this is about leveling the playing field this is about this is about this is the technology and this is what I care about deeply and I know you do too uplifting Humanity enabling every man woman and child to have access to Food Water Healthcare education and have a voice and have a voice they are invisible and have dreams yeah and have dreams and have the

tools to fulfill those dreams to have agency yes agency is the right word I had a bit of a flipping comment because again I can do what I want in my kind of role I was like humanity is creatively constipated we're going to make it so compute rainbows I think that's great it's a silicon but again it's the reality because people don't believe they can create they don't believe they the mentality and mindset is wrong because people have more agents than they can do an individual can shake the world or the individual can make anything around them better but not if they don't believe they can and this is why art therapy is used in mental health settings to amazing things we've been conditioned to consume rather than create yeah we've been conditioned to be polarized rather than talk to each other and communicate with each other and this can again can change that and again that's why like I said this should be in my opinion a human right it is infrastructure as important as 5G and what I'm trying to catalyze now is not that I build the company that makes the decisions for that but that we put it out there and we're spinning off a Lutheran other things I just figure out government structure it's not the UN it's something else so I think one important point is do you think a world in which uh individuals are held back or restricted feel they have no hope or a world where every mother knows their children have access to the best Healthcare the best education you know the best ability to create that's a more peaceful world in my mind yes 100 I mean look all wars are based on lies okay uh because also for otherwise both sides can be couldn't believe it because humans are humans like to kill another human it's disgusting yeah right and so you have to tell the lie that that person is the other and you have to communicate it and control the means of communication you look at kind of again where conflicts are resolved when people realize they are humans and we're all part of a global Society but our infrastructure has been set up to polarize like literally we can see it visually this is how it happens the incentive structure is misaligned so how can you fight polarization if not by communication and how can you do that if you don't give people these tools and you create it so there is a base foundation for the world so there are generalized models that are Global then every country has its own AI policy using their variance those models yeah and then because it's all standardized we can hop between one and the other Nana is a peaceful future yes and and a future worth working towards creating working towards creating but it's also now is the first time we can build that future because of this disruption technology like that you know governments are there's a definition of a government it is the entity with the legitimate use of political violence right the only one that's a that's a sad that's a sad definition but it's the nature of it right because you saw lots of political violence and then it was Consolidated into one entity they can imprison you yes and they've got an army back in the currency and this and that and that right and governments rule on the basis of pure legitimacy to violence and again we see that kind of thing right so against this what typically changes the government or a society it is an act of violence in some ways it's an act of disruption it can be a technology it can be a revolution or anything like that this is a revolution yeah that's happening just after covid when everyone's thinking holy crap the system was rubbish let's do better yeah I I agree and the challenge is that in the world today you can't transform a government gradually yes and this is why as well crypto there were some amazing things in there and amazingly smart people working there it's rubbish because it tried to build a system outside of the existing system and there was this decision system and then there was the interface fortunes were made there and fraud happened there yeah whereas this AI because it can Bridge structured and unstructured can actually go into our systems out compete them and make them more efficient and bring them forward and it's the first technology they can do that dynamically and at scale or build virtualized systems that are de novo that we spend our time in and opt out of the existing system and into the new exactly and they cannot compete oh the

final thing of course is that if you keep it as it is whereby it's controlled by the few they will ultimately use it as a system of control it is the panopticon forever and again we're seeing this in China and other places with a social credit score that's about to be augmented with AI everyone's looking at everyone else monitoring everyone else yeah what is freedom there maybe it's still happiness maybe it's control let's give people the option right let's give people the infrastructure in the building blocks they need to be independent happy and maybe they don't choose Independence maybe it's a bit more kind of book-like that's fine at least you got the choice yeah that's that's because I mean this is again kind of if we tie this all back the fact that you can now have ai that can write better than a human that can draw better than a human that can emote and speak to emotional turns means that let's say for instance take one aspect of it companies that are driven that sell ads they can create the most manipulative ads in the world ever and Regulators will not regulate that that's interesting and they do no but now it's the next step up because they have these latents that resonate so like now when you look at some AI art artists can complain what they want it's resonant when you listen to the most advanced AI voices it's emotional you can feel it it tugs at something and again this is a breakthrough that's literally a year old let's talk about uh the ethics and morals does does AI have a moral compass I think that the creation to have a moral compass well I think the Creator is a very technology is not neutral okay the creative technology do have a responsibility and they will never make it neutral because it embodies their perspective and this bodies the data set and other biases and things like that so I think that AI itself this particular type of AI again if we just take the model it is the action upon the model and then leads to the output so there's a responsibility there but then like how do we adapt them all do we just have a one and done thing that slightly training Western values and norms and mores which is which is the way it's been going historically in the large corporate setting yeah because there's nothing you can do like you can't build a Swahili version of it because you don't have access to technology whereas now with the pre-training and other things like that you can do that with one graphics card that's great right because again we've kind of flipped the Paradigm of and AI needs to be going all the time to this pre-compressed Foundation model so I think that you know then there is the things of like when you've got self-driving cars and other things like that what are the ethical Norms the trolley problem and everything that you input on that these are not easy questions because you're extending Humanity which that means you also extend the ethics of humanity and that is not the same around the world education one of your moonshots uh we first got to know each other through the Global Learning X prize that Elon Musk and Tony Robbins funded uh you were one of the leaders of the one billion team ah that wasn't there I just kind of helped them with Joe idea yes and so uh what so if you don't mind uh what did your team do in that X prize and then where are you taking this Vision so my co-founder Joe and I um Joe LED this made imagine worldwide to take the winners of the x-prize KitKat school and one billion who are the real Chicano Champions and implement it around the world yeah into the rohingya camps and Malawi and camps and others like I just support from the side and cheer him along as he kind of goes and does the really hard valuable work but now we've seen that I believe the later stats from the randomized controlled trials because you need to actually implement it when it happens I think 76 of kids in refugee camps learning literacy and numeracy in 13 months without internet yeah it's immense it's like one hour of use per day is the equivalent of them being in school and pretty much yeah because like as one teach up of 300 students 400 students but then is it enough no it's not enough what we've done at the moment what needs to happen is there needs to be a big Grand Challenge whereby you know Malawi and Malawi kind of has said that they want to roll this out at Super scale and to have

multiple other governments now that we have the rcts let's make an open source ecosystem that has AI at the core that teaches kids and learn from kids so you take from what's Happening Malawi and move it to Ethiopia Sierra Leone Ranger camps Brooklyn everywhere and so there is a actual Superstar amazingly well created ecosystem for education and again go to the Future and bring it back with us love it this year at abundance 360 Sal Khan is going to be joining us as well have you spent any time with with Sal no not yet okay so I'm excited to connect you guys because you know he's built something pretty extraordinary but his vision is to bring AI to it so that it's AI is generating the content not him and it's and it's able to rapidly iterate for cultural appropriateness and so forth this is why we need to build so one of the things is that we're building National level models um from India to other ones where there's localized data sets and other things if you can get the education piece going remember how I said that um this AI is like a kindergartener or a grade schooler sure it matters what you teach it so right now we're teaching it everything do we need to teach everything no yeah so if you're an AI that teaches the kids and learns from the kids that's the best data in the world to teach an AI for a Kenyan model or a Nigerian model or others and you know who should run that technology Nigerians yeah sure and others sure and so one of the things we're doing it's almost like family based learning and extrapolated from there exactly because we don't know best but we can give tools and we've reduced the barriers to create National level localized cultural models and then those models together form a constellation that not only have you got based learning of like what's the optimal way to teach learning algebra getting better then you can go beyond that to that and you can and the plan is to have an integrated system where it's Hardware software deployment curriculum because then we can update that through Mesh networking or the amazing work of project Giga which is putting high-speed internet into every school in the world by the UN and then you can put Healthcare on that yes and then you have a self-adaptive improving Healthcare System self-adaptive improving education system and then the world I mean for me that it is the that's the calling uh that I think both of us have and hopefully many entrepreneurs here it's like what greater purpose could you have in life than uplifting Humanity in that fashion exactly and then as an Etc manager you can fund that at scale by bonds yeah and the world's biggest problems or the world's biggest business opportunities right exactly so I become a billionaire help a billion people so one of the ways that we're kind of doing it is results oriented bonds whereby you can pledge 20 million dollars for million kids are providedly educated on this standardized architecture the invisible become visible and measurable measurable infrastructure banks on the World Bank fund the remainder held by the Pension funds and you can divert billions and billions of dollars into this it's kind of the promise of one laptop a child can finally be done but rather than building a system for today we build a system for tomorrow that constantly adapts and improves is understandable and standardized because that is infrastructure Yes again the thing I really want to emphasize is this ai's infrastructure yeah it's more important than 5G it's oxygen in the room it's oxygen in the room yeah um so when I think about the future of Education going out 10 20 years and bringing it back it today uh for me it's not a book and it's not a flat screen it's going into a virtual world if I want to learn about Plato there's a guy sitting on a slab a marble over there and he says hey Peter come on over let me show you around introducing my friends and it's experiential yes and that that NPC of Plato is trained up by all the knowledge about Plato by every historian and it's accurate and the imagery and so forth and and what you just said earlier about real time rendering from stable diffusion enables that right and the ability to take every historian's work on Plato and train up a model and Plato enables all that exactly and particularly when it is at the hardware level because typically what software happens is that again you build layers

and layers and layers of kind of compilers and translations so you're quite far from the hardware this model is already efficient at the top level what happens when we optimize them and push them down to the hardware level don't need internet you don't need anything you can form it but then all of a sudden you have the young lady's Illustrated primer yes Neil Stevenson's Incredible Book and a vision for for the future of Education exactly yeah and but we can make it finally but we can make it more we can make it closer to the prime Radiance and foundation so are you building Hardware you might getting other people to build it for me my life is we're setting the specs but this is the thing who is we okay the way that we're gonna do it is that is it similar to the Grand challenges and the president like let's get together we will drive the process because the rule by committee you know never works but let's invite everyone to participate from the kids using the tablets to code them yeah to the most advanced developers in the world and let's build something for Humanity that's the way to do this amazing amazing so um you don't like the term web3 you wanted to jump over to uh to web4 but this this virtualized world uh which is the convergence of of AI and uh and uh and VR AR blockchain and so forth what where do you see it in the near term going uh it's just going to go insane so like we have technology uh gonna make an announcement in January about some of our technology you've got Apple likely having AR glasses snap oculuses all of this it'll be a fully immersive worlds where you can engage like when I was a video game investor I looked at time to flow fun and frustration yes and I think this technology can adjust all of those and create worlds for us but it'll be a year or two because again now it's percolating and it's getting ready to then create brand new experiences on the Fly for everyone and I think in a couple of years it starts going in five years creativity imagination engagement entertainment is completely transformed within five years within five years it's going to be the biggest shift change that we've ever seen because the incumbents can't keep up with entities that are using this technology that can do everything like what does making a picture look like when you can change it live with just words right and you can say actually make his hair a bit longer or you know get rid of those nostril hairs and it just understands and does it without having to functionality yes intentionality and action yes you know it's kind of you know the whole military thing of observe oriented decide and act right these systems enable that almost flawlessly like if you use open ai's whisper technology the translation of the podcast is just immense and all these other things because it's learned so many principles and they're tiny files and then you can say you could make it so preferential and say improve it the way you think I'd want it improved you can and like we're building technology for example it tells you how good a story is or how good code is yes so then you have a creation and a discriminator they bounce back and forth against each other yeah and they learn from your personal context amazing you took a break from being a hedge fund manager uh to address your son's autism which was uh I mean there are individuals like yourself like Martin rothblatt and others who are like you know I refuse to believe that I that something can't be done yeah and you jumped in um for those who are dealing with autism uh uh personally or in their families or in their Community what were your learnings and what would you advise there's always room for improvement and everyone struggles with something how old is your son today he's 14 14. get super happy um so I've asked budget is an ADHD myself you know I think they balance each other out that's what they say you're doing pretty well that's all right um so I had a lot of kind of issues around those things um but people again everyone's different and then diversity is Australia as it were but sometimes kids and others need help because they can't achieve their potential because there's too much going on so like I said I buy into the Gaba glutamate balance theory of it and so we had some drug repurposing and other stuff on end of one equals case

that's not science as it were it's first principles and I don't think it's coherent enough to kind of be brought forward but it's interesting instead I think there are certain things that benefit everyone such as applied behavioral analysis whereby because you haven't built up the words there are short trials to try and reconstruct what a cup means it's quite an intensive process but it's also what people do after strokes and other things that make them lose their connectivity in their brain most of this thing though is about noise filtering and reduction of that but I think percolation of that means that you have to look into first principles analysis of some of the science of what can cause it and then you have to bring that forward to what is safe so don't do kind of crackpot theories but there's an emerging science and study showing things like n acetylcysteine sulfur refrain other things that calm you down are probably the best and large part of it is just connection and engagement there so one of the things that we'll be doing next year is that we're taking all the research that I've done and formalizing it properly because again it's not a case of well I can do it so anyone should have it I did an N equals one case yes but then that should be extended out and this is what I realized when covert came along so I designed and launched with the help of loads of people uh Collective and augmented intelligence against covid-19 launched at Stanford in July of 2020. it was that the actual origin of stability that was the first origin of stability so we didn't incorporate at that time but it kind of put it together I mean it's it's insane how far things have gone in two years yeah I mean we probably actually kicked off in 13 months ago um so yeah it was insane because I thought I saw it coming and I saw it like autism as a multi-systemic inflammatory condition where even now if anyone asks you how does covet actually work if you ask a scientist and they'll tell you we don't actually sure sure like why are ferritin levels high and why is this and that the reality is our base Foundation is not good enough we don't have enough shared knowledge so I created that to create a system as comprehensive authoritative and up to date so there's a nice blog about an OCD type site and things like that a lot of private sector Enterprises that promised a lot didn't deliver and so I realized again this technology was the future and we needed to create open infrastructure for that so when we do our autism releases next year all of the knowledge and everything like that will be available there will be a semantic scholar on steroids which allows you to access the information relevant to the type of autism that your child might have because I basically figured out there were 16 different etiologies or driving conditions but thirty percent of kids will get worse and 26 of kids will get better with the same treatment so it doesn't work yeah this area of personalized medicine needs a foundation and again that Foundation must be common but we can't wait around to do it and it's interesting right the more I learn about the fundamentals of human biology the more complicated you can dig a layer after layer after layer yeah and I don't think it's possible for I mean we're discovering so much because the tools were were using but we're going to need this level of uh of AI to cognitively understand the com the interactions of the systems again it's complex hierarchical system yeah you know classical Hub Simone style and so we have to build new tools to enable that the question is are these tools closed and the companies trying to are they open and so I'll take care is open yeah and then so our business models just scale and service around that which is how all the servers and databases are but open is secure as well that's why Linux is used everywhere versus Microsoft Windows Etc so I think the final part of this is that you know again it's all interrelated like the body is such a wonderful powerful thing if you look at longevity and things like that we need again this first principles thinking to both make us healthy and live longer and be happier yeah um have you thought much about what's coming down the pike with Quantum Technologies and Quantum Computing so like I think uh very sympathetic to various approaches that are like I'm quite I quite like the quantum annealing on the kind of d-wave

um because I think that you know Carl Christian's theory around kind of free energy principle and having these low energy states makes a lot of sense in fact this is similar to what the AI models do in that when you look at the latent spaces you're going to the low energy states of what that could kind of mean right um so that's quite a lot of jogging I think for a lot of listeners but basically um Quantum Computing kind of is another part of the puzzle a lot of people try to say one AI model and say oh we're gonna put something it can do everything you know they look at AGR there's a thing that can do everything the human brain is made up of so many different parts and we're just filling in some of the missing gaps right now of which Quantum Computing is one of them yeah and it's it I think it's adding fuel to the fire of how fast things are moving ah yeah it's gonna be and you know it's increasing and we can see it getting there I mean again part of this is a super computer thing right like our super computer we've been the fastest in the world five years ago yeah which is insane for a private company right yes absolutely you know like think about bigger than everything but if you look at like Nvidia is the one that kicked off because they put AI in the core of their graphics cards before this even happened which is why it's now a 32 billion dollar part of their business the av-100s four years ago with the first iteration the a100 so the next now the h100s are like up to nine times faster it's literally going exponential the compute is going exponential the research the Technologies we see as much Technologies everywhere and we're like it's so difficult to piece all these things together see I don't even know where things will be in a year let alone five years later ten years I think that's an important part and this is what when Ray talks about the singularity you know it's the notion that we're you know the ability to predict what's going to happen next has become impossible and that time frame you know people say What's it gonna be like in 20 years I can barely think about 10 years from now or five years from now let alone 20 years yeah and this affects the way that we act so the way that our brains work is that our default is decision making under risk so we look at the upside downside of something and we make an expected utility calculation based on that because systems are stable is the world stable now no yeah you know so what do we do instead we make decisions under uncertainty which is we minimize for maximum regret you know or actually just minimize for regret yeah so with these powerful Technologies people like don't give them out because I don't know what can happen it's the it's the uh it's the amygdala it's a dystopian point of view it's minimizing our downside protecting what we have it's a scarcity and fear-based animal brain you know reptile brain that we default to and this happens synchronicity that's what we saw with covert it just happened all of a sudden everyone thought the same right it's what we're seeing with tech layoffs and things like that yeah like Facebook makes loads of money why are they laying off you know things like that yeah and so we have to be kind of aware of this because what's happening now as well is that the classical system has reached its end we've borrowed too much money from the future and now we are going to a negative sum game with inflation and other things people are losing wealth rapidly that leads to unstable Nash equilibrium from a game theoretic perspective so we go from one steady state and Lurch to another inflation to deflation to maximum employment to job losses to fiscal stuff and it's just going to be a crazy few years and that's the reason I call it stability that was one of my questions word you know and obviously there's there's a terminology of stability in the in the models that you're building but stability in order to help stabilize Society yes build a human Os or catalyze a human OS because it can only be built by Society we're just a capitalist yes help guide that in a way because this AI can be finally the thing that can stabilize a complex system that is humanity and the knowledge to achieve our potential at a platform a infrastructure platform that uplifts all of humanity yes and again it should be run by the people for the people so like our subsidiaries in the countries

we're putting aside 10 of the equity for the kids that use the tablets and I'm never going to take a penny out of any of them untilipoing them because they should be owned by the people I love that okay close this out with two uh two topics uh around the moonshots and mindsets um if I were to launch an uh emod X prize fully funded uh and push it out to the world what isn't what's a Grand Challenge in xprize that you would love to see uh materialize out there I think this education is the key one I think it's the next step of Global Learning which is a Grand Challenge just to build this system and especially for low-income countries again like the Delta on the impact can be so massive on this because it's infrastructure for the Next Generation like a lot of these Emerging Markets leapfrogged straight to mobile phones and Skip computers yeah now they can skip to the AI age how amazing will that be it'll uplift everyone it and and there's no greater asset to a nation a company anybody than the intelligence of its of its citizens citizens are intelligent they don't have access to be able to take that intelligence and build you know to build for themselves and extend that we can make that infrastructure now to do it for the first time ever yeah amazing amazing what are the mindsets that have allowed you to be successful do you think I talk about a abundance mindset a longevity mindset a moonshot mindset an exponential mindset curiosity mindset but any of those resonate for you or are there other mindsets because I think mindsets are the most important differentiator that we have so like I've always been very lucky and I've achieved very interesting things I was never really motivated for the last few years when I finally applied myself and what I'm good at is first principles thinking on this first bits rather than atoms but I realize there's nothing I can't do there's nothing people can do anything yes if they put their mind to it but they have to think in a structured way and they have to let almost water flow as it were so what I always try to do is I try to make it a win-win for everyone to participate to help to extend this and at a time of absolute change you can make that happen um so this is why I go to the Future and I go back to the past and I work back that way which I think a lot of people don't because they get stuck in the present so it's the moonshot mindset looking to the Future and again recursively back propagating and I'll close out with uh if you were going to list the moon shots that you're working on right now you're clearly working in education we've talked about that but as someone who's my disrupt all the industries uh Healthcare as a moonshot I am working on everything everything I've got about 18 different ones but education is cool and creativity is cool those two enable everything else if you want to fix climate if you want to fix hate if you want to fix a lot of different things get those two right and that's the foundation for the future it's a pleasure my friend uh what a fun conversation uh excited to share your wisdom and vision with the world excited for what you're creating as a as a fundamental platform and infrastructure for Humanity uh I wish you all the luck and look forward to supporting you any way I can thank you very much again it's just a little Catalyst it'll be everyone else that drives this forward see you in March at a360 Cheers Cheers everyone this is Peter again before you take off I want to take a moment to just invite you to subscribe to my weekly Tech blog today over 200 000 people received this email twice per week in the tech blog I share with you my insights on converging exponential Technologies what's going on in AI how longevity is transforming adding decades to our life in the tech blog I often look at the 20 meta trends that are going to transform the decade ahead and share the conversations I've had with Incredible Tech thought leaders on how they're transforming Industries if that sounds cool to you and you want to try it join me go to dmads.com backslashblog enter your email and let's start this weekly conversation let me share with you the incredible progress we're seeing in the world of technology and the positive impact it's having on our lives again that's

diamandis.com backslash blog looking forward to sharing my insights an incredible breakthroughs I'm seeing with you every

4. Hugging Face

 Generative AI  Hugging Face

4.1 5 Minutes Generative AI Developer Setup

No this is not a click-bait title. We will use Hugging Face to do all the heavy lifting, however if you have a Mac laptop, you can download a pre-trained deep learning model, perform computer vision inference and visual question answering on a random image from Wikipedia, run a Jupyter Notebook with the demo, all this in 5 minutes flat.

First, you should be running the latest Python on your system with Python package manager upgraded to the latest.

```
python --version
# should return Python 3.10.x or higher as on Jan'23
pip --version
# should return pip 22.3.x or higher as on Jan'23
```

Follow this guide for [Mac OS X](#) if you do not have the latest Python. If installing specific version of Python for managing dependencies then follow [this thread](#) to install using `pyenv` Python version manager. If required upgrade pip to the latest using the following command.

```
pip install --user --upgrade pip
```

We will now create a virtual environment for our MLOps setup so that our dependencies are isolated and do not conflict with the system installed packages. We will follow [this guide](#) for creating and managing the virtual environment. First change to the directory where we will develop our application.

```
python -m venv env
```

If you run `ls env` you will see following folders and files created.

```
bin      include    lib      pyvenv.cfg
```

Now we can activate our virtual environment like so. You will notice that development directory prefixed with the `(env)` to indicate you are now running in the virtual environment.

```
. env/bin/activate
```

You can confirm that you are not running inside the virtual environment with its own Python.

```
which python
## should return /Users/.../env/bin/python
```

To leave the virtual environment using `deactivate` command. Re-enter using same command as earlier.

Now we are ready to install our dependencies for running Hugging Face Transformer models with PyTorch.

```
pip install torch torchvision transformers
```

We can test our installation with the following script.

Test Installation

```
python -c "from transformers import pipeline; \
print(pipeline('sentiment-analysis')('we love you'))"

# [{'label': 'POSITIVE', 'score': 0.9998704195022583}]
```

Now we can setup our development environment which is Jupyter Notebook and [Jupyter Widgets](#).

```
pip install notebook ipywidgets
```

Let us also save our setup in the requirements.txt file.

```
pip freeze > requirements.txt
```

One last thing we will do is to setup a custom cache directory for our models and datasets downloaded when using Hugging Face. Edit your bash script. On zsh shell this is the `~/.zshrc` file. This is required as the default cache directory is in the home directory which is not a good idea as it will fill up your home directory with large files. When you run Hugging Face API for the first time, it will download the models and datasets to the default cache directory. You can change the default cache directory by setting the environment variables.

`~/.zshrc`

```
export TRANSFORMERS_CACHE="/Users/.../cache/transformers"
export HUGGINGFACE_HUB_CACHE="/Users/.../cache/hub"
export HF_HOME="/Users/.../cache/huggingface"
```

That is it! We setup a development environment with Jupyter Notebook. We installed the minimum set of dependencies required to get started developing using deep learning transformer models available at Hugging Face.

We can now test visual question answering transformer model using sample from Hugging Face. Try this in your python shell or a Jupyter Notebook.

Visual Question Answering

```
from transformers import pipeline

vqa = pipeline(model="impira/layoutlm-document-qa")
vqa(
    image="https://huggingface.co/spaces/impira/docquery/resolve/2359223c1837a7587402bda0f2643382a6eefeb/invoice.png",
    question="What is the invoice number?",
)

# [{"score": 0.42514941096305847, 'answer': 'us-001', 'start': 16, 'end': 16}]
```

The above code snippet uses the [LayoutLM](#) model to perform visual question answering on a given invoice image. The model is trained on the [DocQuery](#) dataset.

As on Feb'23, even ChatGPT does not support multimodal question and answering. However, thanks to pre-trained transformer models available at Hugging Face, we could setup a development environment to perform visual question answering in 5 minutes 😊

5. Open AI

 OpenAI  ChatGPT  Generative AI

5.1 OpenAI CEO Sam Altman on AI for the Next Era

We are sharing the video transcript of OpenAI CEO Sam Altman on AI for the Next Era. The video is available on YouTube. We are also demonstrating how to generate meaningful new content using ChatGPT on a long form content like a video transcript. Specifically, we are using the video transcript to generate a summary of the video, a list of key entities, and a Mermaid.js diagram of the relationships between the entities. This is a powerful way to learn about the topics covered in the video. As we see more customer requests for such content generation, we will create more summaries and diagrams for the video transcripts. Generating this content yourself using GPT-3 Davinci model for automation can be a costly proposition (roughly \$150+ for the length of this transcript). We are making this content available to you for free using our paid ChatGPT plus account. Please let us know if you find this useful.

Introduction

Here is the prompt we have engineered to generate the new content for this transcript. We use "as an expert" and "step by step" to indicate that the new content should be written in a step by step manner. We use "based on the following text" to indicate that the new content should be based on the text in the transcript chunk. We use "start generating a summary of the text in your own words" to indicate that the new content should be a summary of

the text in the transcript chunk. We use "use the summary and text to extract and list key entities and types of entities like people, companies, products, technologies, and processes" to indicate that the new content should be a list of key entities and types of entities. We use "describe the relationship between the entities using Mermaid.js diagram notation using a fenced code block" to indicate that the new content should be a Mermaid.js diagram of the relationships between the entities.

ChatGPT Prompt

As an expert in this subject perform these actions step by step based on the following text:

1. Start generating a summary of the text in your own words.
2. Use the summary and text to extract and list key entities and types of entities like people, companies, products, technologies, and processes.
3. Describe the relationship between the entities using Mermaid.js diagram notation using a fenced code block.

Text: <replace with transcript text chunk>

Summary:

The prompt generates following summary of the video transcript. It also generates a list of key entities and a Mermaid.js diagram of the relationships between the entities.

ChatGPT Summary

The text discusses the potential business opportunities that can arise from the use of very large language models through APIs. The author believes that with the quality of language models expected to improve in the coming years, there will be a serious challenge to Google's search product. The author also suggests that human-level chatbot interfaces, new medical services, and education services can become very large companies. The use of language interfaces and powerful models will be a new technological platform, leading to the creation of many new companies. The author suggests that creating an enduring differentiated business is key to success, and a middle layer of startups that tune existing large models may become important.

Key entities

- People: Sam Altman
- Companies: Google, Dolly, co-pilot, startups
- Products: very large language models, APIs, multimodal models
- Technologies: language interfaces, chatbots
- Processes: creating enduring differentiated businesses, tuning large models

```
graph LR
A[Google] --> B[search product]
A --> C[lasting differentiated business]
D[AI businesses] --> C
E[startups] --> F[tuning existing large models]
E --> D
G[very large language models] --> H[APIs]
```

```
G --> I(multimodal models)
J(medical services) --> D
K(education services) --> D
L(language interfaces) --> D
M(powerful models) --> C
```

The above content is generated based on following transcript chunk of roughly 3,500 characters.

ht let's start a little bit more pragmatic but then we'll Branch out so one of the things I think a lot of folks here are interested in is based off the apis that very large models will create what are the real business opportunities like what are the ways to look forward and then how given the the apis will be available to multiple players how do you create distinctive businesses on him yeah um so I I think so far we've been in the realm where it's you know you can do like an incredible copywriting business or you can do like a sort of like Education Service or whatever um but we I don't think we've yet seen the kind of like people go after the like you know trillion dollar like take on Google's um and I think that's about to happen like maybe it'll be successful maybe Google will do it themselves but like I would guess that with the quality of of language models we'll see in the coming years um you know there will be like a serious challenge to Google for the first time for for a search product um and I think people are really starting to think about like how did the fundamental things change um and that's going to be really powerful uh I think that a like a human level chat bot interface that actually works this time around like I I think like you know many of these trends that like we all made fun of were just too early like the chatbot thing was good it was just too early now it can work and I think you know having like new medical services that are done through that where you get great advice or new Education Services like this these are going to be very large companies I think we'll get multimodal models and not that much longer and that'll open up new things I think people are doing amazing work with sort of agents that can use computers to do things for you use programs and this idea of like a language interface um where you know you say a natural language what you want in this kind of like dialogue back and forth you can iterate and refine it and the computer just does it for you you see some of this uh with like Dolly and co-pilot in very early ways but I think this is going to be a massive Trend and you know very large businesses will get built with this as the interface and more generally that like these very powerful models will will be one of the genuine new technological platforms which we haven't really had since mobile and there's always like an explosion of new companies right after so that'll be cool and and what do you what do you things are given that the large language model we provided as an API service what are the things that you think that folks who are thinking about these kind of AI businesses should think about is how do you create an enduring differentiated business so you know they're they're I think there will be a small handful of like fundamental large models out there that other people build on but right now what happens is you know company makes large language model API other people build on top of it and I think there will be a middle layer that becomes really important where uh I'm like skeptical of all of the startups that are trying to sort of train their own models I don't think that's going to keep going but what I think will happen is there'll be a whole new set of startups that take an existing very large model of the future and tune it uh which is not just fine-tuning like all the things you can do I think there will be a lot of access provided to create the model for medicine or using a computer

Data Flywheel Process

ChatGPT Summary

The text discusses the potential value that can be created in a middle layer of companies that create unique versions of language models through a data flywheel process. The author believes that a systemic mistake is to assume that language models will not generate fundamentally new knowledge and add to the sum total of human scientific knowledge. The text also explores how AI can contribute to science, including the use of dedicated science products and tools that increase productivity, as well as the potential for AI to become an AI scientist and self-improve.

Key entities

- People: Kevin Scott
- Companies: AlphaFold, bio companies
- Products: language models, science products, tools
- Technologies: AI, data flywheel process
- Processes: creating unique versions of language models, scientific development, automation of AI developers' jobs

```
graph LR
A(bio companies) --> B(value creation)
C(AI) --> D(self-improvement)
C --> E(automation of AI developers' jobs)
F(science products) --> G(value creation)
H(tools) --> G
```

or like the kind of like friend or whatever and then those companies will create a lot of enduring value because they will have like a special version of they won't have to have created the base model but they will have created something they can use just for themselves or share with others that has this unique data flywheel going that sort of improves over time and all of that so I think there will be a lot of value created in that middle layer and what do you think some of the most surprising ones will be it's a little bit like for example you know a surprise from a couple years ago and we talked a little bit to Kevin Scott about this this morning as we opened up which is train on the internet do code right so so what do you think some of the the surprises will be of you didn't realize it reached that far I think the biggest like systemic mistake in thinking people are making right now is they're like all right you know maybe I was skeptical but this language model thing is really going to work and sure like images video too but but it's not going to be generating net new knowledge for Humanity it's just going to like do what other people have done and you know that's still great that's still like brings the marginal cost of intelligence very low but it's not it's not going to go like create fundamentally new it's not going to cure cancer it's not going to add to the sum total of human scientific knowledge and that is what I think will turn out to be wrong that most surprises the current experts in the field yep so uh let's go to science then there's the next thing what are some of the things whether it's building on the apis you know uh use of apis by scientists where what are some of the

places where science will get accelerated and how so I think there's two things happening now and then a bigger third one later um one is there are these science dedicated products whatever like Alpha fold and those are adding huge amounts of value and you're gonna see in this like like way more and way more I like I think I if I were like you know had time to do something else I would be so excited to like go after a bio company right now like I think you can just do amazing things there um the anyway but there's like another thing that's happening which is like tools that just make us all much more productive uh that help us think of new research directions that sort of write a bunch of our code so you know we can be twice as productive and that impact on like the net output of one engineer a scientist I think will be the surprising way that AI contributes to science that is like outside of the obvious models but even just seeing now like what I think these tools are capable of doing copilot is an example there's you know be much cooler stuff than that um that will be a significant like change to the way that technological development scientific development happens but then so those are the two that I think are like huge now and uh lead to like just an acceleration of progress but then the big thing um that I think people are starting to explore is um I hesitate to use this word because I think there's one one way it's used which is fine and one that is more scary but uh like AI that can start to be like an AI scientist and self-improve and so when like can we automate like can we automate our own jobs as AI developers very first the very first thing we do can that help us like solve the really hard alignment problems

Business Opportunities

ChatGPT Summary

The text discusses the potential business opportunities and impact of language models on various industries, including search, medical services, and education. The importance of the middle layer in the value chain is highlighted, which involves companies using unique versions of language models to create enduring value. The text also covers the potential of AI in science, with tools that make researchers more productive and the development of AI scientists that can automate jobs and help solve complex alignment problems. The alignment problem is defined as the challenge of building AGI that aligns with human goals, and self-improving systems are seen as a potential solution. Finally, the text mentions the evolution of language models as the most certain development in the near future.

Key entities

- People: Kevin Scott
- Companies: Google, OpenAI
- Products: AlphaFold
- Technologies: language models, AI
- Processes: alignment problem

```
graph LR;
    subgraph Language_models
        Language_models --> Enduring_value_companies
```

```

Language_models --> Search_products
Language_models --> Medical_services
Language_models --> Education_services
Language_models --> Language_interface_businesses
Language_models --> Middle_layer
Enduring_value_companies --> Special_version
Middle_layer --> Tunes_existing_models
end
subgraph AI in science
AI_in_science --> Science_dedicated_products
AI_in_science --> Productivity_tools
AI_in_science --> AI_scientists
end
subgraph Alignment problem
Alignment_problem --> AGI
AGI --> Self_improving_systems
Self_improving_systems --> Alignment_research
end
Language_models --> Evolution_of_language_models
AI_in_science --> Evolution_of_AI

```

that we don't know how to solve like that honestly I think is how it's going to happen um the the scary version of self-improvement like the one from the science fiction books is like you know editing your own code and changing your optimization algorithm and whatever else um but there's a less scary version of self-improvement which is like kind of what humans do which is if we try to go off and like discover new science uh you know that's like we come up with explanations we test them we think like we whatever process we do uh that is like special to humans teaching AI to do that I'm very excited to see what that does for the total like I'm a big believer that the only real driver of human progress and economic growth over the long term is the the structure the societal structure that enables scientific progress and then scientific progress itself and uh like I think we're gonna make a lot more of that well especially science that's deployed in technology say a little bit about how what uh I think probably most people understand what the alignment problem is but it's probably worth four sentences on the alignment problem yeah so the alignment problem is like we're going to make this incredibly powerful system and like be really bad if it doesn't do what we want or or if it sort of has you know goals that are uh either in conflict with ours um and many Sci-Fi movies about what happens there or goals where it just like doesn't care about us that much and so the alignment problem is how do we build AGI that that does what is in the best interest of humanity how do we make sure that Humanity gets to determine the you know the future of humanity and how do we avoid both like accidental misuse um like where something goes wrong we didn't intend intentional misuse where like a bad person is like using an AGI for great harm even if that's what other person wants and then the kind of like you know inner alignment problems where like what if this thing just becomes a creature that views this as a threat um the the way that I think the self-improving systems help us is not necessarily by the nature of self-improving but like we have some ideas about how to solve the alignment problem at small scale um and we've you know been able to align open ai's biggest models better than we thought we'd we would at this point so that's good um we have some ideas about what to do next um but we cannot honestly like look anyone in the eye and say we see at 100 years how we're going to solve this problem um but once the AI is good enough that we can ask it to like hey can you help us do alignment research um I think that's going to be a new tool in the toolbox yeah like for example one of the conversations you and I had is could we tell the uh the agent don't be racist right as opposed to trying to figure out all the different things where the weird correlative data that exists on all the training settings everything else may lead to racist outcomes it could actually in fact do a self-cleansing totally once the model gets smart enough that you can that it really understands what racism looks like and how complex that is you can say don't be racist yeah exactly um what do you think are the kind of Moon shots that in terms of evolution of the next couple years that people should be looking out for in terms of like evolution of where AI

we'll go um I'll start with like the higher certainty things I think language models are going to go just much much further

Future of AI

ChatGPT Summary

The speaker is excited about the algorithmic progress to come in the field of AI, including advancements in multimodal models and models that continuously learn. They predict that these advancements will unlock new applications and lead to a genuine technological revolution. They also believe that AI can help with new knowledge generation and advancing humanity. The speaker notes that many people are claiming to use AI in various areas, such as fusion, but these claims may be illusory.

Key entities

- Companies: Open AI
- Products/Technologies: GPT, Transformers
- Processes: algorithmic progress, new paradigm

```
graph LR;
subgraph AI_Advancements
Multimodal_models --> New_applications
Continuously_learning_models --> Advancements_in_AI
New_paradigm --> Future_of_AI
end
subgraph AI_Impact
New_knowledge_generation --> Advancing_humanity
end
subgraph Illusory_AI_Claims
AI_+_Fusion --> Illusory_claims
end
```

than people think and we're like very excited to see what happens there um I think it's like what a lot of people say about you know running out of compute running out of data like that's all true but I think there's so much algorithmic progress to come that that we're going to have like a very exciting time um another thing is I think we will get true multimodal models working and so you know not just text and images but every modality you'd like in one model able to easily like uh you know fluidly move between things um I think we will have models that continuously learn so like right now if you use GPT whatever it's sort of like stuck in time that it was trained and the more you use it it doesn't get any better and all of that I think we'll get that changed so very excited about all of that and if you just think about like what that alone is going to unlock and the sort of applications people will be able to build with that um that that would be like a huge victory for all of us and just like a like a massive step forward and a genuine technological Revolution if that were all that happened um but I think we're likely to keep making research progress into new paradigms as well um we've been like pleasantly surprised on the upside about what seems to be happening and I think uh you know all these questions about like new knowledge generation how do we really Advance Humanity uh I think there will be systems that can help us with that so one thing I think would be useful to share because uh folks don't realize that you're actually making these strong predictions from a fairly critical point of view not just a you know we can take that Hill say a little bit about some

of the areas that you think are current kind of illusionally talked about like for example Ai and fusion oh yeah so I like one of the unfortunate things that's happened is uh you know AI has become like the Mega buzzword um which is usually a really bad sign I hope I hope it doesn't mean like the field is about to fall apart um but historically that's like a very bad sign for you know new startup creation or whatever if everybody is like I'm this with AI and that's definitely happening now um so like a lot of the you know we were talking about like are there all these people saying like I'm doing like these you know RL models for Fusion or whatever and as far as we can tell they're all like much worse than what like you know smart physicists to figure it out um I think it is just an area where people are going to say uh everything is now this plus AI many things will be true I do think this will be like the biggest technological platform of the Generation Um but I think it's like we like to make predictions where we can be on the frontier understand predictably what the scaling laws look like or already have done the research where we can say all right this new thing is going to work and make predictions out from that way and that's sort of like how we try to run open AI um which is you know do the next thing in front of us when we have high confidence and Kate take 10 of the company to just totally go off and explore which has led to huge wins and there will be wait like oh I feel bad to say this like I I doubt we'll still be using the Transformers in five years I hope we're not I hope we find something way better but the transform has obviously been remarkable so I think it's important to always look for like you know where am I going to find the next the sort of the next totally new paradigm um and

Target Markets

ChatGPT Summary

The speaker believes that making predictions through AI is not the only way, as it is essential to observe how something works predictably and gets better. They predict that the cost of intelligence and energy will trend towards zero, causing seismic shifts in society. The application of AI will seep into almost everything, and the cost structure will change. The speaker is uncertain about the metaverse's impact but thinks it has the potential to be a new container for software.

Key entities: AI, Financial markets, Quant trading system, Intelligence, Energy, Metaverse, Computing, Simulation environments, Education, Entertainment

```
graph LR
AI -- applies_to --> Financial_markets
AI -- applies_to --> Quant_trading_system
Intelligence -- trends_towards --> zero
Energy -- trends_towards --> zero
Intelligence -- affects_cost_of --> everything_else
Energy -- affects_cost_of --> everything_else
Metaverse -- may_impact --> Computing
Metaverse -- may_impact --> Simulation_environments
Metaverse -- may_impact --> Education
Metaverse -- may_impact --> Entertainment
Metaverse -- may_be_a --> container_for_software
```

but but I I think like that's the way to make predictions don't don't pay attention to the like AI for everything like you know can I see something working and can I see how it predictably gets better and then of course leave room open for like the you can't plan the greatness but sometimes it had the research breakthrough happens yep so I'm going to uh ask two more questions and then open it up because I want to make sure that people have a chance to

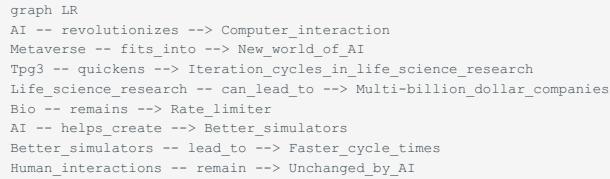
do this uh the broader discussion although I'm trying to paint the broad picture so you can get the crazy aspirations as part of this what do you think uh what do you think is going to happen vis-a-vis the application of AI to like these very important systems like for example financial markets um you know because the very natural thing would be is saying well let's let's do a high frequency Quant trading system on top of this and other kinds of things what what is it is it just kind of being a neutral arms race is it is it what how do how what's your thought in like it's almost like the life 3.0 yeah omega's point of view yeah um I mean I think it is going to just seep in everywhere my basic model of the next decade is that uh the cost of intelligence the marginal cost of intelligence and the marginal cost of energy are going to Trend rapidly towards zero like surprisingly far and and those I think are two of the major inputs into the cost of everything else except the cost of things we want to be expensive the status Goods whatever and and I think you have to assume that's going to touch almost everything um because these like seismic shifts that happen when like the whole cost structure of society change which happened many times before um like the Temptation is always to underestimate those uh so I wouldn't like make a high confidence prediction about anything that doesn't change a lot or that where that doesn't get to be applied um but one of the things that is important is it's not like the thing Trends either Trends all the way to zero they just Trend towards there and so it's like someone will still be willing to spend a huge amount of money on compute and energy they will just get like unimaginable amount of intelligence energy they'll just get unimaginable amounts about that and so like who's going to do that and where is it going to get the weirdest not because the cost comes way down but the amount spent actually goes way up yes the intersection of the two curves yeah you know the thing got 10 or 100 thing got 100 times cheaper in the cost of energy you know 100 million times cheaper in the cost of intelligence and I was still willing to spend a thousand times more in today's dollars like what happens then yep and then uh last of the buzzword Bingo part of the the future questions metaverse and AI what do you what do you see coming in this you know I think they're like both independently cool things it's not like totally clear to me yeah other than like how AI will impact all Computing yeah well obviously Computing simulation environments Asians possibly possibly entertainment certainly education right um you know like an AI tutor and so forth those those would be Baseline but the question is is there anything that's occurred to you that's I I would bet that the metaverse turns out in the upside case then which I think has a reasonable chance of happening the upside case the metaverse turns out to be more like something on the order of the iPhone like a new a new container for software

AI and the Metaverse

ChatGPT Summary

The speaker believes that AI and the metaverse will revolutionize computer interaction, with the metaverse fitting into the new world of AI. They believe foundational technologies such as tpg3 will quicken iteration cycles in life science research, leading to new multi-billion dollar companies. Although some areas of life science research will be impacted by AI, human trials and bio's own pace will remain rate limiters. The speaker believes that AI can help create better simulators, leading to faster cycle times. However, the deep biological things like human interactions will not be changed by AI.

Key entities: AI, Metaverse, Life science research, Pharma companies, Human trials, Bio, Startups, Simulators, Self-improvement



and you know a new way a new computer interaction thing and AI turns out to be something on the order of like a legitimate technological Revolution um and so I think it's more like how the metaverse is going to fit into this like new world of AI then AI fit into the metaverse but low confidence the TBD all right questions hey there how do you see uh Technologies uh foundational Technologies like tpg3 affecting um the pace of life science research specifically you can group in medical research there and and sort of just quickening the iteration cycles and then what do you see as the rate limiter in life science research and sort of where we won't be able to get past because they're just like laws of nature yeah something like that um so I think the currently available models are kind of not good enough to have like made a big impact on the field at least that's what like most like life sciences researchers have told me they've all looked at it and they're like it's a little helpful in some cases um there's been some promising work in genomics but like stuff on a bench top hasn't really impacted it I think that's going to change and I think uh this is one of these areas where there will be these like you know new 100 billion to trillion dollar companies started those those areas are rare but like when you can really change the way that if you can really make like a you know future Pharma company that is just hundreds of times better than what's out there today that's going to be really different um as you mentioned there still will be like the rate limit of like bio has to run at its own thing and human Trials take however long they take and that's so I think an interesting cut of this is like where can you avoid that like where are the the synthetic bio companies that I've seen that have been most interesting are the ones that find a way to like make the cycle time super fast um and that that benefits like an AI That's giving you a lot of good ideas but you've still got to test them which is where things are right now um I'm a huge believer for startups that like the thing you want is low costs and fast cycle times and if you have those you can then compete as a startup against the big incumbents uh and so like I wouldn't go pick like cardiac disease is my first thing to go after right now with like this kind of new kind of company um but you know using bio to

manufacture something that sounds great uh I think the other thing is the simulators are still so bad and if I were an a if I were a bio means AI startup I would certainly try to work on that somehow when do you think the AI Tech will help create itself oh it's almost like a self-improvement will help make the simulators significantly better um people are working on that now uh I I don't know quite how it's going but you know there's very smart people are very optimistic about that yeah other questions and I can keep going on questions I just want to make sure you guys had a chance this uh here yes great Mike is coming awesome thank you um I was curious what what aspects of Life do you think won't be changed by AI um sort of did all of the deep biological things like I think we will still really care about interaction with other people like we'll still have fun and like the reward you know systems of our brain are still going to work the same way like we're still going to have the same like drives to kind of create new things and you know compete for silly status

Video Transcript Continued

and like you know form families and whatever um so I think the the stuff that people cared about 50 000 years ago is more likely to be the stuff that people care about you know 100 years from now than 100 years ago as an amplifier on that before we get to the next whatever the next question is what do you think are the best utopian science fiction universes so far good question um Star Trek is pretty good honestly uh like I do like all of the ones that are sort of like you know we turn our Focus to like exploring and understanding the universe as much as we can um it's not this is not a utopian one well maybe I think the last question is like an incredible short story uh-huh yeah that was what that came to mind yep uh I was expecting you to say Ian Banks on the culture those are great uh I think science fiction is like there's not like one there's not like one sci-fi universe that I could point to and say I think all of this is great but like the collective optimistic corner of sci-fi which is like a smallish Corner um I'm excited about actually uh I took a few days off to write a Sci-Fi story and I had so much fun doing it just about sort of like the optimistic case of AGI um that it made me want to go like read a bunch more so I'm looking for recommendations of more to read now um like the sort of less known stuff if you have anything I will I will get you some great some recommendations so in a similar vein one of my favorite sci-fi books is called childhood's End by Arthur Clark from like the 60s I think and the I guess the one sentence summary is aliens come to the Earth try to save us and they just take our kids and leave everything else so you know there's a slightly more optimistic than that but yes I mean there's Ascension into the over mind is is meant to be more utopian but yes okay uh you may not read it that way but yes well also in our current Universe yes our current situation um you know a lot of people think about family building and fertility and like some of us have different people have different ways of approaching this but from where you stand what do you see as like the most promising Solutions it might not be a technological solution but I'm curious what you think other than everyone having 10 kids you know like how do we of everyone having 10 kids yeah how do you populate how do you like how do you see family building coexisting with you know AGI high tech it's this is like a question that comes up at open AI a lot like how do I think about you know how should one think about having kids there's I think no consensus answer to this um there are people who say yeah I'm not I was gonna I thought I always thought I was gonna have kids and now I'm not going to because of AGI like there's just for all the obvious reasons and I think some less obvious ones there's people who say like well it's going to be the only thing for me to do in you know 15 20 years so of course I'm

going to have a big family like that's what I'm going to spend my time doing you know I'll just like raise great kids and then I think that's what'll bring me fulfillment I think like as always it is a personal decision I get very depressed when people are like I'm not having kids because of AGI uh the EA Community is like I'm not doing that because they're all going to die they're kind of like a techno optimists are like well it's just like you know I want to like merge into the AGI and go off exploring the universe and it's going to be so wonderful and you know just I want total freedom but I think like all of those I find quite depressing um I think having a lot of kids is great I you know want to do that now more than I did even more than I did when I was younger and I I'm excited for it what do you think will be the way that most users interact with Foundation models in five years do you think there'll be a number of verticalized AI startups that essentially have adapted and fine-tuned Foundation models to an industry or do you think prompt engineering will be something many organizations have as an in-house function I don't think we'll still be doing prompt Engineering in five years I think it'll just be like you and this will be integrated everywhere but you will just like you know either with text or voice depending on the context you you will just like interface in language and get the computer to do whatever you want and uh that will you know apply to like generate an image where maybe we still do a little bit of prompt engineering but you know it's kind of just going to get it to like go off and do this research for me and do this complicated thing or just like you know be my therapist and help me figure out how to make my life better or like you know go use my computer for me and do this thing or or any number of other things but I think the fundamental interface will be natural language let me actually push on that a little bit before we get to the next question which is I mean to some degree just like we have a wide range of human talents right now uh and taking a look for example a dolly when you have like a great visual thinker they can get a lot more out of Dolly because they know how to think more they know how to iterate the loop through the test don't you think that will be a general truth about most of these things so it isn't that why would be natural language is the way you're doing it it will be there will be like almost an evolving set of human talents about going that extra mile 100 I just hope it's not like figuring out to like hack the prompt by adding one magic word to the end that like changes everything else I I like what will matter is like the quality of ideas and the understanding of what you want so the artist will still do the best with image generation but not because they figured out to like add this one magic word at the end of it because they were just able to like articulate it with a creative eye that you know I don't have certainly what they have is a vision and kind of how their visual thinking and iterating through it yeah yeah well obviously it'll be that word or prompt now but it'll iterate to to better all right uh at least we have a question here hey thanks so much um I think the term AGI is used uh thrown around a lot and um sometimes I've noticed my own discussions like the sources of confusion has just come from people having different definitions of AGI and so it can kind of be the magic box where everyone just kind of projects their ideas onto it and I just want to get a sense from you what like how do you think you know how would you define AGI and how do you think you'll know yeah it's a great point I think there's like a lot of valid definitions to this but uh for me um AGI is basically the equivalent of a median human that you could like you know hire as a co-worker um so and then they could like say do anything that you'd be happy with a remote co-worker doing like just behind a computer which includes like you know learning how to go be a doctor learn how to go be a very competent coder like there's a lot of stuff that a media human is capable of getting good at and I think one of the skills of an AGI is not any particular Milestone but the The Meta skill of learning to figure things out and that it can go decide to get good at whatever you need um so for me like that's that's kind of like AGI and

then Super intelligence is when it's like smarter than all of humanity put together thanks um just uh what would you say or in the next 20 30 years are some of the main societal issues that will arise as AI continues to grow and what can we do today to mitigate those issues obviously the economic impacts are huge and I think it's just like if it if it is as Divergent as I think it could be for like some people doing incredibly well and others not I think Society just won't tolerate at this time and so figuring out when we're gonna like disrupt so much of economic activity and even if it's not all disrupted by 20 or 30 years from now I think it'll be clear that it's all going to be um what like what is the new social contract like how do my guess is that the things that we'll have to figure out are how we think about fairly Distributing wealth um access to AGI systems which will be like kind of the commodity of the realm and governance like how we collectively decide what they can do what they don't do things like that um and I think figuring out the answer to those questions is is gonna just be huge I I'm optimistic that people will figure out how to spend their time and be very fulfilled I think people worry about that in a little bit of a silly way I'm sure what people do will be very different but we always solve this problem um but I do think like the concept of wealth and access and governance those are all going to change and how how we address those will will be huge actually one thing I don't know what level of devs you can share that but one of the things I love about what openai and you guys are doing is when you they think about these questions a lot themselves and they initiate some research so you've initiated some research on this stuff yeah so we run the largest uh Ubi experiment in the world I don't think that is uh we have a year and a half a year and a quarter left in a five-year project I don't think that's like the only solution but I think it's a great thing to to be doing um and you know I think like we should have like 10 more things like that that we try um we also try with different ways to get sort of input from a lot of the groups that we think will be most affected and see how we can do that early in the cycle um we've explored more recently like how this technology can be used used for reskilling people that are going to be impacted early um we'll try to do a lot more stuff like that too yeah so they are the the organization is actually in fact uh these are great questions addressing them and actually doing a bunch of interesting research on it so next question hi so um creativity came up today in several of the panels you know and um it seems to me that the way it's being used like you you have tools for human creators to go and expand human creativity so where do you think the line is between these tools to to allow a Creator to be more productive in artificial creativity itself so um I I think and I think we're seeing this now that tools for creatives that that is going to be like the great application of AI in the short term um people love it it's really helpful uh and I think it is at least in what we're seeing so far um not replacing it is mostly enhancing it's replacing in some cases uh but for the majority of like the kind of work that people in these fields want to be doing it's enhancing and I think we'll see that Trend continue for a long time um eventually yeah it probably is just like you know we look at 100 years okay it can do the whole creative job um I think it's interesting that if you asked people 10 years ago uh about Holly I was going to have an impact with a lot of confidence from almost most people you would have heard you know first it's going to come for the blue collar jobs working in the factories truck drivers whatever then it will come for the kind of like the low skill White Collar jobs then the very high skill like really high IQ uh white-collar jobs like a programmer or whatever and then very last of all and maybe never it's gonna take the creative jobs and it's really gone exactly and it's going exactly the other direction and I think this like isn't there's an interesting reminder in here generally about how hard predictions are but more specifically about you know we're not always very aware maybe even ourselves of like what skills are hard and easy like what uses most of our brain and what doesn't or how like difficult bodies are to

control or make or whatever we have one more question over here hey thanks for being here so you mentioned that um you will be skeptical of any startup trying to train their own language model and it would love to understand more so what I have heard and which might be wrong is that large language models depend on data and compute and any startup can access to the same amount of data because it's just like internet data and compute like different companies might have different compute but I guess I see a big players can sell more compute so how good a large language model startup differentiate from another how would the startup differentiate from another how would one large language model startup differentiate I think it'll be this middle layer um I think in some sense the startups will train their own models just not from the beginning uh they will take like you know base models that are are like hugely trained with a gigantic amount of compute and data and then they will train on top of those to create you know the model for each vertical and and that those startups so in some sense they are training their own models just not not from scratch but they're doing the one percent of training that really matters for for whatever this use case is going to be those startups I think they will be hugely successful and very differentiated startups there but that'll be about the kind of like data flywheel that the startup is able to do the kind of like all of the pieces on top of and Below uh like this could include prompt engineering for a while or whatever the sort of the kind of like core base model I think that's just going to get too too complex and too expensive and the world also just doesn't make enough chips so Sam has a work thing he needs to get to so and as you probably can tell with a very far far ranging thing Sam always expands my uh boundaries and a little bit unlike the that when you're feeling depressed whether it's kids in a house you're the person I always turn to probably I appreciate that yes so anyway I think I think like no one knows like we're sitting on this like precipice of AI and like people like it's either gonna be like really great or really terrible um you may as well like you gotta you gotta like plan for the worst you certainly like it's not a strategy to say it's all going to be okay but you may as well like emotionally feel like we're going to get to the Great future and we'll play as hard as you can to get there and play for it yes rather than like act from this place of like fear and despair all the time because if we acted from a place of fear and paranoia we would not be where we are today so let's thank Sam for spending dinner with us thank you

6. Models

6.1 Large Language Models

We are creating a reference of popular and recently released large language models. This is a work in progress and we welcome contributions. Our objective is to demonstrate PromptxAI API capabilities in generating human-readable content which is also structured and factually accurate for programmatic access via an API.

Why Large Language Models Matter

GPT-3 davinci prompt

```
As an expert in the field of machine learning
explain, in a narrative style, large language models with
factually accurate information using anecdotes from industry
influencers and thought leaders.
```

Large language models are a type of artificial intelligence (AI) that use deep learning techniques to process natural language. They are being used in a wide range of applications, from chatbots to summarizing text. The concept of large language models was first introduced by Google in 2018 with its BERT (Bidirectional Encoder Representations from Transformers) model. BERT was trained on a large corpus of text to learn the relationships between words and phrases. This enabled it to better understand the context of the text and produce more accurate results.

Since then, the use of large language models has grown exponentially. According to one industry influencer, Andrew Ng, “Large language models are revolutionizing natural language processing. They are more accurate, faster, and easier to use than traditional language models.” Large language models are also being used to generate text. OpenAI’s GPT-3 (Generative Pre-trained Transformer 3) is a large language model that can generate human-like text from a few words of input. It is being used to generate articles, stories, and even code.

The potential of large language models is immense. As Andrew Ng said, “The ability to generate human-like text is a game-changer for natural language processing. It opens up a whole new world of possibilities for AI applications.” Large language models are also being used to improve the accuracy of machine translation. Google’s Translatotron is a large language model that can translate from one language to another without relying on a separate machine translation system. This has the potential to significantly reduce the time and cost associated with traditional machine translation systems. The possibilities of large language models are exciting, and the technology is only going to get better. As another industry influencer, Fei-Fei Li, said, “Large language models are going to be the foundation of the next wave of AI applications.”

List of Large Language Models

GPT-3 davinci prompt

Only respond using markdown with accurate facts from reputable sources.
 Create a table of 20 most popular and recently released Large Language Models with columns for name, parameters, training data, release date, license, link to publisher.

Name	Parameters	Training Data	Release Date	License	Link to Publisher
GPT-3	175 billion	Common Crawl, BooksCorpus, WebText	June 2020	OpenAI API	OpenAI
T5	11 billion	C4, Wikipedia, BooksCorpus, WebText	May 2020	Apache 2.0	Google AI
BERT	340 million	BooksCorpus, Wikipedia	October 2018	Apache 2.0	Google AI
XLNet	560 million	BooksCorpus, Wikipedia	June 2019	Apache 2.0	Google AI
RoBERTa	355 million	BooksCorpus, Wikipedia	October 2019	Apache 2.0	Facebook AI
ALBERT	18 million	BooksCorpus, Wikipedia	October 2019	Apache 2.0	Google AI
ELECTRA	125 million	BooksCorpus, Wikipedia	March 2020	Apache 2.0	Google AI
BART	400 million	C4, BooksCorpus, Wikipedia	May 2020	Apache 2.0	Facebook AI
Reformer	1.6 billion	BooksCorpus, Wikipedia	June 2020	Apache 2.0	Google AI
Longformer	1.6 billion	BooksCorpus, Wikipedia	June 2020	Apache 2.0	AI2
XLM-R	550 million	BooksCorpus, Wikipedia	June 2020	Apache 2.0	Facebook AI
CTRL	1.6 billion	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Salesforce Research
TAPAS	1.6 billion	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Google AI
MT-DNN	1.6 billion	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Microsoft Research
DeBERTa	355 million	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Microsoft Research
SpanBERT	355 million	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Microsoft Research
UniLM	1.6 billion	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Microsoft Research
ERNIE 2.0	1.6 billion	BooksCorpus, Wikipedia	August 2020	Apache 2.0	Baidu Research
Megatron-LM	8.3 billion	BooksCorpus, Wikipedia	August 2020	Apache 2.0	NVIDIA
XLM	550 million	BooksCorpus, Wikipedia	September 2019	Apache 2.0	Facebook AI
XLM-RoBERTa	550 million	BooksCorpus, Wikipedia	April 2020	Apache 2.0	Facebook AI

 **Image generation**

6.2 DALL·E Mini

Generate images based on text prompts for research and personal consumption

Model Provider	Model License	Model Version	Model Release
Hugging Face	Unknown	DALL·E Mini and DALL·E Mega	June 9, 2022

Model Summary

This model card focuses on the DALL·E Mini model, which is used to generate images based on text prompts for research and personal consumption. Intended uses include supporting creativity, creating humorous content, and providing generations for people curious about the model's behavior. The model was trained on unfiltered data from the Internet, limited to pictures with English descriptions, and the model developers discuss the limitations of the model further in the DALL·E Mini technical report. The model developers used 3 datasets for the model and the model is 27 times smaller than the original DALL·E and was trained on a single TPU v3-8 for only 3 days. DALL·E Mega is still training and has been training for about 40-45 days on a TPU v3-256. The model should not be used to intentionally create or disseminate images that create hostile or alienating environments for people, and using the model to generate content that is cruel to individuals is a misuse of this model.

Model Resources

 [Hugging Face](#) |  [GitHub](#) |  [Website](#)

Info

This model card was generated using [PromptxAI API](#) querying recent web content sources with large language model generations. As of Feb 2023 it is not possible to query models like GPT-3 (via applications like ChatGPT) on the latest web content. This is because the model is trained on a static dataset and is not updated with new web content. PromptxAI API solves this problem by chaining recent web content sources with large language model outputs. This allows you to query models like GPT-3 on latest web content.

Model Details

Task: Generating images based on text prompts

Model Parameters: 27 times smaller than the original DALL·E and was trained on a single TPU v3-8 for only 3 days

Model Training Data: 2 million images for fine-tuning the image encoder, 15 million images for training the Seq2Seq model

Model Evaluation Data: Unknown

Model Hyperparameters: Unknown

Model Training Procedure: Images and descriptions pass through the system, fine-tuning the image encoder, training the Seq2Seq model

Model Evaluation Procedure: Comparing DALL·E Mini's results with DALL·E-pytorch, OpenAI's DALL·E, and models consisting of a generator coupled with the CLIP neural network model

Model Strengths: Generate images based on text prompts for research and personal consumption, supporting creativity, creating humorous content, and providing generations for people curious about the model's behavior

Model Limitations: White and Western culture asserted as a default, and the model's ability to generate content using non-English prompts is observably lower quality than prompts in English, may also reinforce or exacerbate societal biases

Model Unique Features: 27 times smaller than the original DALL·E and was trained on a single TPU v3-8 for only 3 days

Model Comparison with Similar Models: DALL·E-pytorch, OpenAI's DALL·E, and models consisting of a generator coupled with the CLIP neural network model

Model Use Cases: Generate images based on text prompts for research and personal consumption, supporting creativity, creating humorous content, and providing generations for people curious about the model's behavior

Model Compute Infrastructure Required: TPU v3-8 and TPU v3-256

 Language Models for Dialog Applications

6.3 LaMDA

Built by fine-tuning a family of Transformer-based neural language models specialized for dialog

Model Provider	Model License	Model Version	Model Release
Google	Unknown	Unknown	Unknown

Model Summary

This post describes the LaMDA project, which is a family of Transformer-based neural language models specialized for dialog. LaMDA has three key objectives: Quality, Safety, and Groundedness, which are measured using carefully designed metrics. The model is trained in two stages: pre-training and fine-tuning. Pre-training is done on a dataset of 1.56T words, while fine-tuning is done on a dialog dataset restricted to back-and-forth dialog between two authors. Quality metrics generally improve with the number of model parameters, with or without fine-tuning, while Safety does not seem to benefit from model scaling alone, but does improve with fine-tuning. Groundedness improves as model size increases, but fine-tuning allows the model to access external knowledge sources and effectively shift some of the load of remembering knowledge to an external knowledge source. LaMDA's level of Sensibleness, Specificity and Interestingness unlocks new avenues for understanding the benefits and risks of open-ended dialog agents. Exploring new ways to improve the Safety metric and LaMDA's groundedness is the main focus going forward.

Model Resources

 [Github](#) |  [Website](#) |  [About Model](#)

Info

This model card was generated using [PromptxAI API](#) querying recent web content sources with large language model generations. As of Feb 2023 it is not possible to query models like GPT-3 (via applications like ChatGPT) on the latest web content. This is because the model is trained on a static dataset and is not updated with new web content. PromptxAI API solves this problem by chaining recent web content sources with large language model outputs. This allows you to query models like GPT-3 on latest web content.

Model Details

Task: Generate natural-language responses to given contexts, and classification tasks on whether a response is safe and high-quality

Model Parameters: 137B

Model Training Data: 1.56T words from public dialog data and other public web documents

Model Evaluation Data: Multi-turn two-author dialogs

Model Hyperparameters: Unknown

Model Training Procedure: Pre-train the model using GSPMD to predict every next token in a sentence, given the previous tokens; fine-tune the model to perform a mix of generative tasks to generate natural-language responses to given contexts, and classification tasks on whether a response is safe and high-quality

Model Evaluation Procedure: Collect responses from the pre-trained model, fine-tuned model, and human raters; ask a different set of human raters a series of questions to evaluate these responses against the Quality, Safety, and Groundedness metrics.

Model Strengths: Quality metrics (Sensibleness, Specificity, and Interestingness) generally improve with the number of model parameters, with or without fine-tuning

Model Limitations: Below human levels in safety and groundedness

Model Unique Features: Leverages external knowledge sources

Model Comparison with Similar Models: Quality gap to human levels can be narrowed, though the model's performance remains below human levels in safety and groundedness

Model Use Cases: Natural language processing research, program synthesis, zero-shot learning, style transfer, as well as in the BIG-bench workshop

Model Compute Infrastructure Required: Unknown

 Language model  transformer architecture

6.4 LLaMA

Auto-regressive language model, based on the transformer architecture

Model Provider	Model License	Model Version	Model Release
FAIR team of Meta AI	Non-commercial bespoke license	Version 1	December 2022 - Feb 2023

Model Summary

Meta AI's FAIR team developed the auto-regressive language model LLaMA between December 2022 and February 2023. It is based on the transformer architecture and comes in different sizes, ranging from 7B to 65B parameters. It is intended for research on large language models, such as exploring potential applications, understanding capabilities and limitations, and evaluating and mitigating biases, risks, and toxic and harmful content generations. The primary intended users are researchers in natural language processing, machine learning, and artificial intelligence. The model was trained on data from the Web, and thus reflects biases from this source. It was evaluated on RAI datasets to measure biases exhibited by the model, as well as on eight standard common sense reasoning benchmarks. The data used to train the model contains offensive, harmful, and biased content, and thus the model is not intended to inform decisions about matters central to human life. Risks and harms of large language models include the generation of harmful, offensive, or biased content, and incorrect information. As a foundational model, it should not be used for downstream applications without further investigation and mitigations of risks.

Model Resources

 [Github](#) |  [Paper](#) |  [About Model](#)

Info

This model card was generated using [PromptxAI API](#) querying recent web content sources with large language model generations. As of Feb 2023 it is not possible to query models like GPT-3 (via applications like ChatGPT) on the latest web content. This is because the model is trained on a static dataset and is not updated with new web content. PromptxAI API solves this problem by chaining recent web content sources with large language model outputs. This allows you to query models like GPT-3 on latest web content.

Model Details

Task: Research on large language models

Model Parameters: 7B, 13B, 33B and 65B parameters

Model Training Data: CCNet [67%], C4 [15%], GitHub [4.5%], Wikipedia [4.5%], Books [4.5%], ArXiv [2.5%], Stack Exchange[2%]

Model Evaluation Data: BoolQ, PIQA, SIQA, HellaSwag, WinoGrande, ARC, OpenBookQA, NaturalQuestions, TriviaQA, RACE, MMLU, BIG-bench hard, GSM8k, RealToxicityPrompts, WinoGender, Crows-Pairs

Model Hyperparameters: Table I - Summary of LLama Model Hyperparameters

Model Training Procedure: Kneser-Ney language model and a fastText linear classifier

Model Evaluation Procedure: Evaluated on RAI datasets to measure biases exhibited by the model for gender, religion, race, sexual orientation, age, nationality, disability, physical appearance and socio-economic status, measure the toxicity of model generations, depending on the toxicity of the context used to prompt the model, evaluated on eight standard common sense reasoning benchmarks.

Model

Strengths: Can be used for research on large language models, including exploring potential applications, understanding capabilities and limitations of current language models, and developing techniques to improve those

Model Limitations: Generates incorrect information, prone to generating toxic or offensive content

Model Unique Features: Auto-regressive language model, based on the transformer architecture

Model Comparison with Similar Models: Not applicable

Model Use Cases: Research on large language models, including exploring potential applications such as question answering, natural language understanding or reading comprehension, understanding capabilities and limitations of current language models, and developing techniques to improve those, evaluating and mitigating biases, risks, toxic and harmful content generations, hallucinations

Model Compute Infrastructure Required: Not specified

 Masked language modeling  MLM  English language  Transformers model

6.5 BERT base model (uncased)

Transformers model pretrained on a large corpus of English data in a self-supervised fashion

Publisher	License	Version	Release
Hugging Face team	Unknown	Unknown	Unknown

Model Summary

BERT is a transformers model pretrained on a large corpus of English data in a self-supervised fashion. It was introduced in a paper and first released in a repository. It is uncased, meaning it does not differentiate between English and english. It can be used for masked language modeling or next sentence prediction, but is primarily intended to be fine-tuned on downstream tasks such as sequence classification, token classification or question answering. It was trained on 4 cloud TPUs for one million steps with a batch size of 256 and a sequence length of 128 tokens for 90% of the steps and 512 for the remaining 10%. When fine-tuned on downstream tasks, it achieves good results on Glue test results.

Model Resources

 [Hugging Face](#) |  [Github](#) |  [Research Paper](#)

Info

This model card was generated using [PromptxAI API](#) querying recent web content sources with large language model generations. As of Feb 2023 it is not possible to query models like GPT-3 (via applications like ChatGPT) on the latest web content. This is because the model is trained on a static dataset and is not updated with new web content. PromptxAI API solves this problem by chaining recent web content sources with large language model outputs. This allows you to query models like GPT-3 on latest web content.

Model Details

Size: Unknown

Use Cases: Sequence classification, token classification or question answering

Training corpus: BookCorpus, English Wikipedia

Training method: Adam optimizer, learning rate of 1e-4, **$\beta_1=0.9$, $\beta_2=0.999$** , weight decay of 0.01, learning rate warmup for 10,000 steps and linear decay of the learning rate after

Evaluation method: Glue test results

Compute: 4 cloud TPUs in Pod configuration (16 TPU chips total)

Features: Lowercased and tokenized using WordPiece and a vocabulary size of 30,000

Limitations: Primarily aimed at being fine-tuned on tasks that use the whole sentence (potentially masked) to make decisions

Strengths: Can use lots of publicly available data

6.6 Generative Pre-trained Transformer 3 (GPT-3)

Autoregressive language model released in 2020 that uses deep learning to produce human-like text

Publisher	License	Version	Release
OpenAI	Microsoft has exclusive licensing of GPT-3 for Microsoft's products and services	GPT-3.5	May 28, 2020

Model Summary

GPT-3 is a deep learning language model released in 2020 that uses generative pre-training to produce human-like text. It is the third-generation language prediction model in the GPT series, successor to GPT-2 created by OpenAI. GPT-3 has a capacity of 175 billion parameters and is capable of performing zero-shot, few-shot and one-shot learning. It can generate text that is difficult to distinguish from human-written text, and has potential applications in both beneficial and harmful applications. Microsoft has exclusive licensing of GPT-3, while OpenAI provides a public-facing API. There are concerns about the environmental impact of training and storing the model, as well as potential issues with academic integrity.

Model Resources

 [Research Paper](#) |  [GitHub](#) |  [Wikipedia](#)

Info

This model card was generated using [PromptxAI API](#) querying recent web content sources with large language model generations. As of Feb 2023 it is not possible to query models like GPT-3 (via applications like ChatGPT) on the latest web content. This is because the model is trained on a static dataset and is not updated with new web content. PromptxAI API solves this problem by chaining recent web content sources with large language model outputs. This allows you to query models like GPT-3 on latest web content.

Model Details

Size: 175 billion parameters, requiring 800GB to store

Use Cases: Text generation, summarizing texts, answering questions, coding in CSS, JSX, and Python

Training corpus: Common Crawl, WebText2, Books1, Books2, Wikipedia

Training method: Generative pre-training

Evaluation method: Human evaluators

Compute: 800GB

Features: Zero-shot, few-shot and one-shot learning, edit and insert capabilities

Limitations: Potential to perpetuate fake news, environmental impact of training and storing the models, potential for misuse

Strengths: Eerily good at writing amazingly coherent text, improved language understanding performances in natural language processing (NLP)



6.7 Flan 20B

Instruction tuned with Flan, receptive field of 2048, no mode tokens, trained on C4 corpus

Publisher	License	Version	Release
Google AI	Apache	20B	Q2 2022

Model Summary

This text describes the release of a new open source Flan 20B model that was trained on top of the already open sourced UL2 20B checkpoint. It has the same configuration as the original UL2 20B model, except that it has been instruction tuned with Flan. It is expected to improve the usability of the original UL2 model and has been released on Apache license. The text also discusses the relative improvements of Flan-UL2 20B compared to other models in the Flan series, as well as the limitations of Flan-style models. Finally, it is noted that the release of Flan-UL2 20B expands the size ceiling of the current Flan-T5 models by approximately 2x.

Model Resources

[Hugging Face](#) | [Research Paper](#) | [Demo](#) | [About Model](#)

Model Details

Size: 19.5B parameters

Use Cases: N-shot prompting, few-shot in-context learning

Training corpus: C4 corpus

Training method: UL2 objective

Evaluation method: Big-Bench hard and MMLU

Compute: 7-8 times faster than Flan-PaLM 62B

Features: Expands size ceiling of Flan-T5 models by 2x

Limitations: Instruction tuned on primarily academic tasks, not ideal for open ended generation

Strengths: Best open source model at the moment on Big-Bench hard and MMLU

7. Tags

Following is a list of relevant tags:

7.1 Analytics Function

- [Natural Language Analytics and Visualization](#)

7.2 Apache

- [Flan 20B](#)

7.3 ChatGPT

- [Natural Language Analytics and Visualization](#)
- [OpenAI CEO Sam Altman on AI for the Next Era](#)

7.4 ChatGPT Model

- [Generating Diagrams from World Knowledge](#)
- [Historical Fact Question Answer](#)

7.5 Consulting Industry

- [Generating Diagrams from World Knowledge](#)

7.6 Creative Function

- [Cinematic Alien Art Generation](#)

7.7 Dataset Generation

- [Historical Fact Question Answer](#)

7.8 Deep learning

- [Generative Pre-trained Transformer 3 \(GPT-3\)](#)

7.9 Diagram Generation

- Generating Diagrams from World Knowledge

7.10 Education Industry

- Historical Fact Question Answer

7.11 English language

- BERT base model (uncased)

7.12 Film Industry

- Cinematic Alien Art Generation

7.13 Flan

- Flan 20B

7.14 Generative AI

- PromptxAI API
- OpenAI CEO Sam Altman on AI for the Next Era
- 5 Minutes Generative AI Developer Setup

7.15 Generative pre-training

- Generative Pre-trained Transformer 3 (GPT-3)

7.16 Google AI

- Flan 20B

7.17 Hugging Face

- PromptxAI API
- 5 Minutes Generative AI Developer Setup

7.18  **Image Generation**

- [Cinematic Alien Art Generation](#)

7.19  **Image generation**

- [DALL-E Mini](#)

7.20  **Instruction tuning**

- [Flan 20B](#)

7.21  **Knowledge Management**

- [Generating Diagrams from World Knowledge](#)

7.22  **Language Models for Dialog Applications**

- [LaMDA](#)

7.23  **Language model**

- [LLaMA](#)

7.24  **MLM**

- [BERT base model \(uncased\)](#)

7.25  **Masked language modeling**

- [BERT base model \(uncased\)](#)

7.26  **Midjourney Model**

- [Cinematic Alien Art Generation](#)

7.27  **Natural language processing (NLP)**

- [Generative Pre-trained Transformer 3 \(GPT-3\)](#)

7.28  **OpenAI**

- [PromptxAI API](#)
- [OpenAI CEO Sam Altman on AI for the Next Era](#)

7.29  **Research Function**

- [Historical Fact Question Answer](#)

7.30  **Stability AI**

- [PromptxAI API](#)

7.31  **T5**

- [Flan 20B](#)

7.32  **Transformer**

- [Generative Pre-trained Transformer 3 \(GPT-3\)](#)

7.33  **Transformers model**

- [BERT base model \(uncased\)](#)

7.34  **UL2**

- [Flan 20B](#)

7.35  **transformer architecture**

- [LLaMA](#)