



Santa Clara Investment Fund Research Report

Technology Sector

Fall 2023



Table of Contents

1. Introduction and Foreword	2
2. Artificial Intelligence Deep Dive	4
2A. High Level Overview of AI	4
2B. Large Language Models (LLMs)	6
2C. Artificial General Intelligence (AGI)	8
3. AI Competitive and Landscape Analysis	12
3A. Key Players:	12
3B. OpenAI's Pivot to Profit and Microsoft Ownership:	15
4. Necessary Inputs for AI Technology	16
4A. Computing Hardware (CPUs, GPUs, TPUs)	16
4B. Cloud Computing and Hyperscalers	16
4C. Datasets	18
5. Technology Industry Briefs	20
5A. Financial Technology (Fintech)	20
5B. Computing Hardware	24
5C. Consumer Technology	27
5D. e-Commerce	29
5E. Software as a Service (SaaS)	32
6. Sources	35



1. Introduction and Foreword

About SCIF and Foreword:

The Santa Clara Investment Fund is the premier student-run investment fund at Santa Clara University. We take pride in our holistic approach in understanding the market which enables authentic insights and actionable investment beliefs. Our organization comprises of amazing students from every school within Santa Clara University at each grade level. If you are reading this as a current student, alumni, or even an industry professional, there are countless ways that we believe you could contribute. To find more information on how to get involved, please visit santaclarainvestmentfund.com.

As of writing this in Fall 2023, AI has taken the economy and public interest by storm. Huge breakthroughs in this area continue to happen day-by-day, some of which are already disrupting entire industries at an unprecedented pace. While AI can certainly be seen as the vast unknown, we believe this is an exciting opportunity for individuals to gain knowledge and help position themselves for the future.

In this report, the SCIF TMT (Technology, Media, and Telecommunications) team has broken down every major facet of AI into basic first principles to build a foundational understanding. Included in this report are our analyses behind model training, the competitive landscape, and the auxiliary inputs needed for AI to exist. Additionally, we have given our brief industry stances for select Technology sub-sectors and how AI may impact them. Because this report contains differing levels of nuance for each section, we hope there will be valuable takeaways for any reader.

I cannot thank my team enough for the hard work they have put into this report. Please feel free to check out any of their profiles (LinkedIn's are hyperlinked) on the next page. We are curious to hear any subsequent feedback or thoughts!

— Brendan Plunkett, TMT Director



Technology Team:



Brendan Plunkett

TMT Director

Class of 2024

Ryan Abbott

TMT Senior Advisor

Class of 2024



Morgan Cushing

Analyst

Class of 2024

Elliott Kob

Analyst

Class of 2025

Lola Lewis

Analyst

Class of 2025



Enow-Ebot Akem

Analyst

Class of 2025

Zack Ferreira

Analyst

Class of 2025

Henry Oliviera

Analyst

Class of 2027

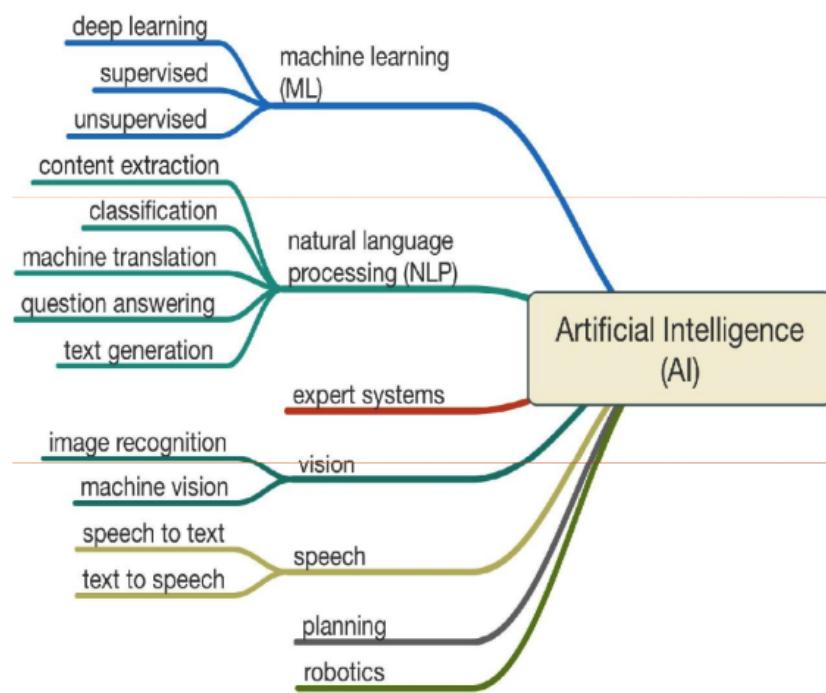


2. Artificial Intelligence Deep Dive

2A. High Level Overview of AI

Understanding AI:

Artificial Intelligence (AI) is a branch of computer science that allows computers to perform tasks that have traditionally been performed by humans. Large Language Models (LLMs), perhaps the most widely recognized form of AI by the public, typically utilize neural networks, are trained on vast amounts of data, and are capable of tasks such as text generation, question answering, translation, and summarization. However, AI encompasses much more than just LLMs or natural language processing; other branches of AI include machine learning, vision, speech, planning, and robotics. Fundamentally, AI models/systems analyze data to draw conclusions and perform tasks.



Deep Learning and Neural Networks:

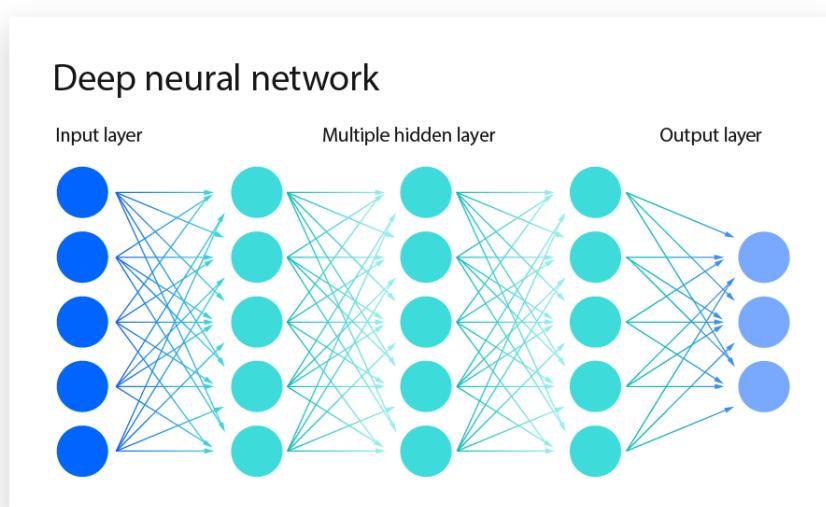
Leading the rapid advancement of AI in the past decade has been global data accumulation. As the world transitions into a more digital realm, the amount of data available for an AI to learn on has grown exponentially. Deep learning is a set of algorithms that utilize data and identify patterns in order to improve. This is possible because of neural networks; as more data becomes available and tasks become more complicated, neural networks increase



the number of layers and interconnected nodes allowing the AI to process and analyze significant amounts of data in a short amount of time.

Key Terms:

- **Neural Network:** An algorithm that utilizes data to recognize and analyze patterns.
- **Deep Learning:** Falling under the category of machine learning (ML), DL uses multi-layered neural networks. Its "depth" is the number of layers between input and output nodes.
- **Training Data:** The input an algorithm utilizes while attempting to complete a task.
- **Compute:** The capacity of processors and the efficiency of running an algorithm.
- **Moore's Law:** The number of transistors in an integrated circuit will double about every two years.
- **GPU:** Graphics Processing Unit. Designed to improve the processing of output (images & videos) on a screen.
- **TFLOPS:** Tera Floating Point Operations Per Second. A performance metric for a computer.
- **Parallelism (in computing):** The simultaneous execution of processes or tasks.
- **Batch Size:** The number of training examples utilized in one iteration.
- **Neural Net Operations:** Operations carried out during the training or operation of neural networks.
- **FLOPS:** Floating Point Operations Per Second. A measure of a computer's performance.
- **Algorithmic Parallelism:** Techniques that allow algorithms to be executed in parallel.





Examples of industry impact in various sectors:

- **Healthcare:** Analyzing and synthesizing vast amounts of patient data with potential diagnosis. Decreases a doctor's cognitive load by optimizing for certain outcomes.
- **Finance:** Answer questions, make predictions, analyze data, fraud detection, etc.
- **Marketing & Sales:** Organization, predictive analysis, automate outreach, analyze research, and trends.
- **Supply Chain and Manufacturing:** AI streamlines operations, predicts maintenance needs, and optimizes the supply chain. Automates operations, tracks maintenance needs, improves efficiency and reduces the likelihood of error.
- **Security & Surveillance:** Facial recognition systems, predictive policing, and digital data security are increasingly AI-driven.
- **Energy Management:** Can actively assess necessary energy outputs, predictive activation, and variable response.

Outlook on AI at-a-glance:

AI is likely to continue growing primarily due to technological improvement, increased available data, and further use cases. Within the next few years, AI can be expected to create jobs and perform tasks that are currently unthinkable. However, at present, the opportunity with the highest probability of success is the advancement of well-established business with the integration of AI. AI will reduce costs through increased efficiency, reduced headcount, and creative problem-solving by analyzing previously untapped data. Adoption rates will only increase in the coming years.

2B. Large Language Models (LLMs)

High-level LLM Overview:

A language model is a type of artificial intelligence model that predicts the most likely word to appear given a sequence of words. A large language model (LLM) is a type of language model that is trained to solve common language problems. ChatGPT is a prominent LLM that can solve these common language problems, examples of which include question answering, text summarization, and text generation. LLMs like ChatGPT can understand and generate human-like text on a scale we've never seen before. This cuts down the time spent on menial tasks and allows us to interpret and analyze text, information, and data much more efficiently, leading to the streamlining of real-world processes, and ultimately, better work.



ChatGPT is very user-friendly, there is a message box that the user uses to give the model a prompt. These models are meant to emulate humans, i.e. they are conversational, so prompts can be things like: “what is cloud computing,” or “write me a short story,” or a summary of text you provide.

LLMs are called “large” for two reasons: first, their ginormous training dataset, which can contain upwards of 1 million gigabytes (a petabyte) of data, and second, the large number of parameters, which is the knowledge that the LLM gained from the training. These parameters define the skill level of each LLM.

How LLMs Are Trained:

LLMs learn to predict the next word in a sentence, which is the basis for things like question answering, through training. Using the dataset provided, the model is given a sequence of words and is asked to predict the next word that will follow. For example, the model could be given the sequence “the sky is...” and then asked to predict the next word. The model begins by making a random guess, and through many iterations of this guessing and checking process, it refines its parameters until it has “learned” the correct answer (i.e., it can reliably and accurately predict and generate coherent sentences). This process is known as self-supervised training. This training is rather easy to do, as the only necessary step for developers is to feed the model a large dataset. There is no need for “data labeling” where the model would need to be shown what the correct answer is, because it simply checks the original dataset to see if its guess matches with the answer.

These models need to be “fine-tuned” in order to become capable of solving niche problems in specific fields. For reference, ChatGPT’s free version uses their GPT-3.5 model, which is a general-purpose LLM, meaning it’s not designed for a specific use case. So, if we want to turn the model into something that can be used for a specific application, we need to fine-tune it. Fine-tuning can be done in three different ways, but the easiest to understand is self-supervised fine-tuning - the same training as mentioned before. The second way is supervised training, which has an input-output style dataset. Essentially, the model will learn to map the input to the output, meaning it specifically associates one input (e.g. a question) with the preset, correct output (the answer to the question). The third way is via reinforcement learning, which starts with the supervised learning from above, then uses a reward model to score each set of responses given to each prompt - the developers show the model which responses score high and low. Finally, it reinforces this learning by passing the prompt through the supervised fine-tuned model and then passing the response through the reward model, which then gives feedback to the fine-tuned model.

The difference between the original self-supervised training and this self-supervised fine-tuning lies in the data fed to the model. For example, if you want a model to write like



Shakespeare, you would feed the model all of Shakespeare's work, and then, with self-supervised training, it would do this guess-and-check process. Once fine-tuned, these models are considered to have domain knowledge. In our Shakespeare case, the model would now have Shakespearean domain knowledge.

Commercial Purposes:

There are use cases for LLMs in nearly all industries you can think of. Fine-tuned LLMs power the chatbots that countless companies use across multiple fields. Tech companies can use chatbots to provide customer support, consumer companies can tailor messages to individual customers, entertainment companies and/or content creators can use models to generate emails, social media posts, videos, and more, and the medical industry can provide information to patients for self-management - the list goes on. An example of this is Duolingo, which uses ChatGPT to power two features they offer: AI conversations in a foreign language and explanations about grammatical rules when a mistake is made. As this technology continues to rapidly develop, more use cases and commercial purposes will undoubtedly arise.

2C. Artificial General Intelligence (AGI)

AGI Overview:

Artificial General Intelligence (AGI) is a hypothetical form of article intelligence in which a machine could develop the ability to comprehend, reason, and interpret information similarly to human intelligence. The goal of the technology is to operate independently of human interference, allowing it to analyze complex problems and come to its own conclusions creatively. There is not an official definition of AGI, and many experts in the field disagree with each other on their understanding of this technology. However, the central aspects of AGI revolve around seven key areas:

1. Knowledge Representation & Reasoning
2. Learning and Adaptation
3. Multi-model learning & Interference
4. Planning & Decision-Making
5. Natural Language Understanding
6. Natural Language Generation
7. Cognitive Architectures



It is arguable if there are current examples of artificial general intelligence. Some technologies that may be considered to have early-stage examples of AGI include driving automation, supercomputers, medical imaging technologies, chatbots, and humanoid robots.

Artificial Intelligence vs. Artificial General Intelligence:

Artificial intelligence, otherwise known as weak AI, is often programmed to perform a task in a specific manner and rely on algorithms or pre-programmed rules to guide their actions. It will correlate questions and assignments with specific datasets to accomplish that project. As a result, AI is limited to operating in certain environments due to the precise methods implemented in its programming.

Artificial general intelligence or strong AI, implements reasoning into its problem-solving process and adapts to new environments rather than just applying an algorithm or coding process. Its flexibility allows it to handle new situations in different industries or sectors.

Advancing Artificial General Intelligence:

There have been many different procedures used to assess the progress of artificial general intelligence. A few of the most famous ones are mentioned below:

The Turing Test: This test was originally proposed by British mathematician and computer scientist Alan Turing to test a machine's ability to be indistinguishable from the likes of a human. It involves a one-hour interrogation where a college-educated panel asks a multitude of questions, examining dialect, logic, reasoning, and other behaviors to determine whether they are engaging with a human or a computer. If a majority of the interrogators cannot distinguish between the computer and the person, then the machine is considered an intelligent entity. The Turing Test has become less reliable over the years as technology has become smarter, and many computer programs have easily passed it using conversational tricks. The test is subjectively based on the opinion of each individual panelist, and each one may have varying standards of human-like behaviors.

The Online University Student Test: The artificial general intelligence participates in an online university and must obtain a college degree. It would also need to communicate with professors and students to mimic the activities of an individual student. The Online University Student test presents some of the same drawbacks as the Turing Test such as developing conversational tricks and subjectivity.

The Artificial Scientist/Nobel Prize Test: The artificial general intelligence must conduct high-quality and original scientific research, including selecting a research problem, performing



research from credited sources, and writing and publishing a paper. The paper must meet the scientific community's expectations such that it could be in contention for a Nobel Prize. One major challenge with this test to consider is the possibility of information bias, resulting in biased or incomplete research that would impact the validity of its findings.

While these evaluations offer some insights into the progression of AGI, there are many others used. These evaluations also do not necessarily provide major improvements to the technology. More research and development in corresponding fields must be done in order to see significant progress in AGI. These areas include:

- **Data Growth:** Publicly available data plays a crucial role in training AI models, and it's a valuable resource for both researchers and organizations working on various AI applications. It relies on data to automate repetitive learning and discovery processes. AI systems leverage large volumes of data to identify patterns, make predictions, and discover valuable insights. Two main drivers of data growth in the future will be proprietary data and research data. Proprietary data is unique to an organization, giving a competitive edge. It can train models to understand business processes, customer behavior, and industry-specific nuances, resulting in more accurate and relevant predictions and recommendations. Research data would be gathered from information being gathered in fields of testing in AGI. An example of research data would be placing robots that resemble humans and have basic functions among people on a large scale to gather a lot of data that imitates our senses. This data can improve the current state of technology by creating a better training feedback loop.
- **Deep Learning:** There is a need to develop new algorithms, techniques, and architectures that can enable AGI to learn, reason, and adapt in a way that is similar to human intelligence. For example, AGI systems should be capable of learning and reasoning not just from isolated data points, but by recognizing patterns, relationships, and abstract concepts. Developing algorithms that can discover context and build abstractions is a critical next step. AGI should also develop problem-solving and reasoning skills through advanced reasoning mechanisms, such as logic, probabilistic reasoning, and causal reasoning. AGI should not just be intelligent but aware of ethical standards and social intelligence, interacting with humans in a way that respects norms and values.
- **Natural Language Understanding:** Humans record and transmit skills and knowledge through books, articles, blog posts, and tutorial videos. As AI systems strive to emulate



human learning and problem-solving capabilities, they must develop the capacity to not only ingest but also comprehensively understand these sources of information. However, this also brings forth challenges related to bias, misinformation, and ethical considerations that must be addressed as AI systems advance in this capacity.

- **Creativity:** In order to accomplish this self-improvement, AI systems will have to rewrite their own code, requiring AI systems to understand the vast amounts of code that humans put together, and identify novel methods for improving it. AI systems aiming for self-improvement need a deep understanding of the vast array of code written by humans. This includes software applications, algorithms, libraries, and frameworks. Understanding code means not only parsing and interpreting it but also grasping the underlying logic, purpose, and potential areas for improvement. It might involve techniques such as natural language processing and code analysis to gain insights from comments, documentation, and code structure.
- **Quantum Computing:** The potential of quantum computing is predicted to have an enormous impact on society. Its proposed use is to apply it for highly complex statistical problems that current computing power cannot address. Since quantum computing can process a vast number of possibilities simultaneously, it can be utilized to speed up AI algorithms and process larger data sets more efficiently, creating a more powerful AI model. By being able to handle vast amounts of data, quantum computing can also help train LLMs, reducing error in its decision-making. However, the quantum computing hardware and software to handle problems required for advancements in AI may not arrive until 2035 or later.
- **Embodied Cognition:** Robots will need to learn from their surroundings through a diverse array of sensory inputs, such as humans do in their developmental phases. They will need to experience the physical world through a vessel similar to a human body, facilitating cognitive advancement in a manner akin to the human experience.



3. AI Competitive and Landscape Analysis

3A. Key Players:

Microsoft & OpenAI (GPT):

Microsoft began its partnership with OpenAI back in 2016 and has pushed its Azure service as the underlying infrastructure for training OpenAI's GPT model. Upon the release of ChatGPT's demo in 2022, its technology went viral on social media and has become the most popular AI chatbot tool today. On March 13, 2023 OpenAI launched ChatGPT-4, granting user's access to the full extent of the model's capabilities with a paid subscription.

OpenAI's most known product line is the GPT (Generative Pre-trained Transformer) series. The company released GPT-2 in 2019 as a simple text generator model to assist in making predictions on what the user might input next when writing. GPT-2 was composed of 1.5 billion parameters and was trained on a data set of 8 million web pages. In June 2020, OpenAI released ChatGPT-3, a state-of-the-art language model with 1.75 billion parameters. To make that possible, Microsoft designed a supercomputer for OpenAI consisting of 285,000 CPU cores and 10,000 GPUs. ChatGPT-3's main uses are for answering questions and providing information on a broad range of topics, drafting various types of texts, offering creative content, text summarization, and more. OpenAI made ChatGPT-3 accessible through an API, allowing developers and businesses to integrate it into applications and services. In its latest version, the model has been upgraded to ChatGPT-3.5, including fewer parameters and improving its machine learning algorithms. OpenAI's most recent model ChatGPT-4 was launched on March 14, 2023. Unlike its predecessors, it can accept both text and image inputs to generate human-like responses. For example, you could submit a graph and the model would be able to present its analysis based on the submission. ChatGPT-4 possesses a staggering 1.76 trillion parameters and uses public and third-party seller's data to increase its sophistication. ChatGPT-4 is the only model in the series that must be paid for via subscription.

OpenAI's other major product offering is the DALL-E series, a learning model designed to generate images given textual descriptions. It contains 12 billion parameters and was trained with text and image pairs. One major limitation of the original DALL-E was that it could only produce images with basic and cartoonish features. However, the series' second model DALL-E2 produces more realistic images with higher resolution. It can also edit parts of an image or photograph based on instructions given, providing a seamless result. The most recent model DALLE-3 understands nuance and detail more than its previous configurations, creating more accurate images based on a description. This was achieved by training an image captioner to generate detailed textual descriptions for images. DALLE-3 is available for free when using Bing Chat or Bing Image Generator.



Artificial intelligence researchers believe LLMs like ChatGPT-4 and Google's PaLM are a part of a new wave of models that exemplify more intelligence than previous models. These researchers have been running experiments with ChatGPT-4, which was trained using an immense scale of compute and data. They explored its capabilities in a vast range of topics such as coding, medicine, law, psychology, and more. Testing was also conducted in activities like playing games, using tools, and explaining itself. From these experiments, the researchers claim ChatGPT-4 demonstrated human-like behaviors such as reasoning, creativity, and deduction. For further development, researchers point toward developing missing components of AGI into LLMs and expand research in understanding the origin of intelligence exhibited in ChatGPT-4.

Google (LaMDA & PaLM 2):

Google, in the last few years has developed its own LLM developed by the company's research team. The models are named LaMDA and PaLM 2. Both LaMDA and PaLM 2 were built on Transformer, Google's neural network architecture that it invented and open-sourced in 2017. By making the bold choice to develop its own LLM, Google differentiated itself from other popular chatbots such as ChatGPT and Bing Chat, which both outsource a language model in the GPT series.

In March, Google released its own chatbot Google Bard to compete with the likes of ChatGPT. However, the release of Bard's demo turned out to be a disastrous launch, with the chatbot infamously delivering inaccurate information about the James Webb Space Telescope. Along with a poor release, many users found Bard to be less capable and conversational compared to ChatGPT. Things only worsened for Bard in September when it was discovered that conversations with Bard appeared in search results. Google has confirmed this indexing was unintentional and is working to block Bard chat transcripts from appearing in search results.

Despite suffering many setbacks, Google continues to make positive steps toward developing Bard. With the update of Bard Extension, customers can use the tool to find relevant information from other Google tools such as Gmail, Docs, Drive, Google Maps, YouTube, Google Flights, and hotels even when the information is across multiple apps and services. Bard's "Google it" function will read the response and evaluate whether there is content across the web to substantiate it. Bard has created another advantage for itself by offering an integrated system for users, whereas ChatGPT is a stand-alone service besides the use of ChatGPTZero.

It will be interesting to see whether or not Google's chatbot will be able to catch up with ChatGPT. Google has a lot of financial and technical resources at their disposal to invest in R&D to continue and develop their AI technology. Additionally, Google possesses one of the largest cloud computing infrastructures in the world. This gives the company a strong foundation to build upon with their current AI models. Although ChatGPT has garnered lots of attention over the past year, Google has a massive audience through Google Search, which has captured 86%



of the search market share. The extent of Google's resources provide continuous, seamless improvement and development that could possibly allow the company to catch up with ChatGPT in the near future.

Meta (LLaMa):

Meta released its foundational LLM LLaMa in 2023 to assist researchers in advancing their work in AI. Its model is considered much easier to train due its smaller size, requiring less computing power and less resources for testing than ChatGPT. LLaMa was trained with 1.4 trillion tokens from publicly available sources. Meta released the original model with the intention of keeping it open-source to the AI community to run on less expensive hardware, developing an upper-hand against ChatGPT on price point. LLaMa's initial release included four models: 7B, 13B, 33B, 65B parameters. According to most natural language processing standards, LLaMa's 13B model surpassed the performance of GPT-3 on most benchmarks despite it being 10x smaller. When tested on its ability to answer human-like questions, LLaMa apparently also outperformed ChatGPT and PaLM in almost all benchmarks and code generation in standard benchmarks.

In July, Meta released LLaMa 2 freely to developers, start-ups, and others interested in creating custom variations of the model. LLaMa 2 was trained with 2 trillion tokens from publicly available sources and offers 40% more data than the original LLaMa. It is also free for research and commercial use, staying dedicated to the belief in open-access models. LLaMa 2 comes in the form of a chatbot called LLaMa-2 chat. Additionally, Microsoft and Meta announced a partnership that provides access to LLaMa 2 on Microsoft's Azure and Windows. While LLaMa 2 outperforms many other open-source models with 70B parameters, it is still not considered as efficient as ChatGPT-4 with 1.76T parameters.

Meta's vice president of generative AI Ahmad Al-Dahle stated Meta is already working on LLaMa 3, but has not gone into detail on its differentiators.

IBM (Watson):

IBM Watson is a supercomputer capable of answering questions posed in natural language, developed in IBM's DeepQA project. With rising interests in AI, demand for the product is demonstrated by IBM's earnings, with Q4 2022 results indicating that Data & AI revenues grew 4% and automation rose 9%. IBM was ranked #1 by IDC for AI lifecycle software market share in February 2022, and the company proclaims that 70% of global banks and 13 of the top 14 systems integrators use Watson.

Watson's services are being applied to various sectors including customer service, supply chain, financial planning, risk and compliance, advertising, IT, video and security at scale. General Motors uses Watson Assistant to automate chatbot responses. Watson was also



embedded into SAP solutions to provide AI-driven insights and automation to accelerate innovation and create more efficient and effective user experiences across the SAP application portfolio.

3B. OpenAI's Pivot to Profit and Microsoft Ownership:

Pivot to Profit:

OpenAI, the artificial intelligence research laboratory, created a for-profit organization in 2019. This decision was made after realizing that donations alone (which is what they were previously operating on) would not scale with the increasing costs of their continuous research and development. OpenAI's overarching power still comes from their non-profit (controlled by the board of directors), which is the majority owner of their capped-profit company.

Microsoft OpenAI Deal Structure:

In 2019, Microsoft invested a total of \$1 billion in OpenAI (with \$500 million in the form of Azure credits), which made Microsoft the exclusive cloud computing services provider to OpenAI. Their cloud computing service is known as Microsoft Azure, or Azure for short, which competes with Amazon Web Services and Google Cloud. Azure is extremely powerful and will aid OpenAI in rapidly developing their model technology. Between 2019 and 2021, they invested an additional \$2 billion. In early 2023, Microsoft invested another \$10 billion, bringing its total investment to \$13 billion. After funding from a few more backers, OpenAI's valuation was brought to \$29 billion. This deal gave Microsoft an exclusive commercial partnership to integrate, distribute, and develop new products on top of OpenAI's technology. The structure of Microsoft and OpenAI's deal is quite unique and important to understand. Microsoft receives 75% of profits until they get back their \$13 billion and from then on have the right to 49% of profits until \$92 billion is earned. Once this mark is hit, Microsoft's shares will go back to OpenAI's non-profit. This is why it is known as a "capped-profit" structure. As of late September 2023, OpenAI is in talks with investors about allowing employees to sell their shares, which would bring their valuation between \$80 billion to \$90 billion.



4. Necessary Inputs for AI Technology

4A. Computing Hardware (CPUs, GPUs, TPUs)

Central Processing Units (CPUs):

CPUs function as the computer's brain, handling inputs, storing vital information, and generating outputs. They're indispensable for anyone using a computer, and their speed significantly influences overall performance. CPUs are available in various configurations, featuring different numbers of cores like single, dual, quad, and hexa cores, catering to diverse computing requirements.

Graphic Processing Units (GPUs):

GPUs are specialized processors crafted to accelerate graphics rendering and process multiple inputs concurrently. Their importance shines in applications such as machine learning, video editing, and gaming, where parallel processing is key. GPUs can be integrated into a computer's structure or added as distinct components.

Tensor Processing Units (TPUs):

TPUs pioneered by Google, are tailored for managing extensive data, especially mathematical operations crucial for machine learning algorithms. Although not as robust as CPUs or GPUs, TPUs excel in processing neural networks, allowing computers to progressively learn and process information much like human brains. Their operation involves streaming data into a queue, processing it within temporary high bandwidth memory, completing computations, and then storing the outcomes. Various TPU versions, such as TPU v2, v3, and v4, offer differing levels of efficiency and speed, addressing a broad spectrum of computational needs. These specialized processors are vital for individuals and organizations engaged in training intricate models, where the training period often spans weeks or even months.

4B. Cloud Computing and Hyperscalers

Hyperscalers:

Hyperscalers operate massive data centers or cloud organizations. They have the ability to "hyperscale" in terms of infrastructure to support more data/computing. Cloud Service Providers (CSPs) like Microsoft Azure, Amazon Web Services, Google Cloud are all examples of cloud/data organizations. Other than cloud services, large tech companies also operate data



centers. The largest being Meta and Apple for their various services: Meta for its social media platforms, Apple for their services like Apple Music, Maps, Stores, Pay, etc.

The actual infrastructure remains very simple using modular premade servers that have embedded CPUs, TPUs, and GPUs. This makes it easier to do maintenance and rapidly expand to computing needs. There is something to be noted about the energy efficiency of hyperscalers, some are much more efficient than others. Below a number closer to 1 is the most optimal, you can see who leads in the sustainability efforts for data centers.

Data Center Operator	Power Usage Effectiveness (PUE)
Meta Platforms	1.09
Google Cloud	1.10
Microsoft Azure	1.22
Alibaba Cloud	1.25
Data Center - Average	1.55

Hyperscalers that are used for AI applications like machine learning, big data analytics, AR/VR tech, demand more power than the average data center pulling in higher bills and higher input costs (more expensive GPUs). All hyperscalers data centers require immense and advanced cooling systems. Infrastructure requires large amounts of space, servers, cables and racks to hold everything, networking to support the servers, and cooling. This process, if done in house, is estimated to take 12-24 months, cost ~\$215M, and take up ~165,000 square feet.



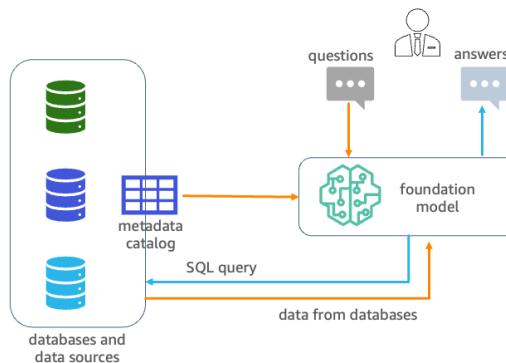
4C. Datasets

Dataset Overview:

Datasets serve as the lifeblood of AI systems, providing the necessary inputs that drive machine learning algorithms.

Architecture Pattern : Answering questions against SQL-based data sources

- Ask questions about factual information without requiring the user to have knowledge of SQL database.
- The foundation model will generate an appropriate SQL query based on the user's question and available metadata.
- The generated SQL query is executed against the database, and the results are returned to the user through the foundation model.



One of the primary functions of datasets in AI is training machine learning models. AI systems learn from data by identifying patterns, relationships, and trends within the data. Datasets enable the model to generalize from the examples it has seen during training. The larger and more diverse the dataset, the more robust and accurate the AI model becomes. Without a substantial and representative dataset, AI models would lack the knowledge and context needed to make meaningful predictions or decisions.

Dataset Impact on AI Training:

Datasets are essential for helping an AI learn to make predictions. Supervised learning, a dominant machine learning paradigm, relies on labeled datasets where each data point is associated with a corresponding target or label. These labeled datasets are essential for training AI models to recognize and make predictions based on specific patterns. For example, in image recognition, labeled datasets containing images and their corresponding object labels are crucial for training AI to recognize objects in new images. On the other hand, unsupervised learning algorithms, such as clustering and dimensionality reduction, benefit from datasets without explicit labels. These algorithms discover hidden structures and relationships within data. Datasets serve as the foundation for unsupervised learning, enabling AI systems to identify patterns, group similar data points, and extract meaningful insights, even when human labels are absent. Transfer learning, a powerful AI technique, leverages pre-trained models on



vast and diverse datasets. These models have learned valuable features and representations from extensive data, making them adaptable to new tasks with relatively smaller datasets. Transfer learning reduces the data requirements for training specific AI applications, thereby increasing efficiency and scalability. Datasets are not only essential for training AI models but also for evaluating their performance. Separate datasets are used for validation and testing to ensure that AI systems generalize well to new, unseen data. The quality of these validation and test datasets directly influences the accuracy, reliability, and safety of AI applications.

Datasets also play a critical role in addressing issues related to bias and fairness in AI. Biased datasets can lead to biased AI models, resulting in unfair or discriminatory outcomes. Careful curation and preprocessing of datasets, along with ongoing monitoring, are essential to mitigate biases and ensure AI systems treat all individuals fairly and equally. In practical AI applications, such as natural language processing, autonomous vehicles, and healthcare, datasets from the real world are invaluable. Real-world data provides AI systems with exposure to the complexity and variability of the environments they will operate in, enabling them to adapt and make informed decisions.

In short, datasets serve as the necessary inputs that fuel the learning, decision-making, and predictive capabilities of AI systems. Their importance cannot be overstated, as they underpin the entire AI development process. Datasets enable AI models to generalize, adapt, and make meaningful contributions across a wide range of applications. As AI continues to advance, the careful selection, curation, and ethical handling of datasets will remain paramount to ensuring that AI systems are accurate, fair, and beneficial to society.



5. Technology Industry Briefs

5A. Financial Technology (Fintech)

Industry Brief

Fintech refers to the integration of technology into offerings by financial services companies to improve their use and delivery to consumers. Companies in the finance industry that use fintech have expanded financial inclusion and use technology that are convenient, efficient, and cost-effective alternatives to traditional financial institutions and methods. Some major areas of Fintech include:

1. **Payment and Money Transfer Services-** Fintech technology offers various payment and money transfer solutions, including mobile wallets, peer-to-peer payment apps, and digital payment platforms, creating a seamless process for sending and receiving money.
2. **Online Lending and Crowdfunding-** Fintech firms provide online lending platforms, peer-to-peer lending, and crowdfunding services that connect borrowers with lenders or investors, simplifying the borrowing and lending process.
3. **Robo-Advisors-** These are automated, algorithm-based platforms that provide investment advice and portfolio management services, typically at lower costs than traditional financial advisors.
4. **Anti-Fraud-** Fintech plays a crucial role in helping prevent and detect fraudulent activities in the financial sector. Fintech companies leverage advanced technologies and data analytics to enhance security measures and protect against various forms of fraud.
5. **Personal Finance-** Fintech apps and platforms help individuals manage their finances by providing budgeting tools, expense tracking, and investment planning.
6. **Blockchain and Cryptocurrency-** Fintech plays a significant role in the development and utilization of blockchain technology and cryptocurrencies like



Bitcoin and Ethereum. Blockchain technology offers secure, transparent, and decentralized systems for various financial applications, including digital currencies and smart contracts.

Key Players

Key players within the Fintech industry include Square, Paypal, Visa, Mastercard, Affirm, Intuit, Robinhood, and SOFI.

AI Case Studies Within Key Players

Mastercard

Mastercard delved into the AI space in 2017 when they acquired Brighterion. Brighterion provides real-time artificial intelligence technology to 74/100 of the largest U.S. banks and more than 2,000 companies worldwide. Brighterion and Mastecard can implement this AI on such a large scale due its distributed architecture and patented, self-learning technology.

AI and machine learning secures more than 100 billion transactions annually to help leading organizations manage the credit risk lifecycle and predict delinquency, prevent payments and acquirer fraud, detect healthcare fraud, waste and abuse, and more. Mastercard has been leveraging its AI capabilities and its unique network view of account-to-account payments, it is helping banks predict and prevent payments to scams of all types. Mastercard has collaborated with nine prominent UK banks, such as Lloyds Bank, Halifax, Bank of Scotland, NatWest, Monzo, and TSB, to leverage extensive payment data for the proactive detection of payment fraud as it occurs, preventing the withdrawal of funds from a victim's account. Through the application of Mastercard's AI-driven cybersecurity solutions, they have successfully thwarted more than \$35 billion in fraudulent losses over the past three years.

Visa

One of the ways Visa provides value-added services with AI includes RTP PreventService, or instant payment services. Instant payment services refer to a type of electronic fund transfer or payment method that allows for the immediate transfer of money from one bank or financial institution to another. Through the technology, Visa can increase security and optimize efficiency. The RTP PreventService provides a multi-FI risk analysis feed that augments existing risk processes and enhances transaction risk management. This allows Visa to support real-time payment authorization and monitor suspicious fraudulent behaviors.

At the beginning of October, Visa announced a \$100 million generative AI ventures initiative to invest in the next generation of companies focused on developing generative AI technologies and applications that will impact the future of commerce and payments. This initiative will be led by Visa Ventures, the global corporate investment arm of Visa.



Block Inc.

Similarly to peers, The company's system allows a company to analyze every transaction, looking for patterns and indicators that could indicate potential fraudulent activity. Machine learning algorithms constantly learn from extensive historical data, improving their ability to differentiate between legitimate and fraudulent transactions over time. Block's approach greatly reduces the necessity for manual intervention in identifying and reviewing transactions, resulting in a more streamlined fraud prevention process that optimizes resource allocation within the organization.

AI Impact

Developing fintech alongside AI requires a strategic approach that takes into account emerging technologies, changing consumer needs, and evolving regulatory landscapes. Here are some of the potential ways that AI and broader technological advancements will impact the industry:

- **Quantum and Edge Computing-** Quantum computing represents a groundbreaking leap in computational power by harnessing the principles of quantum mechanics. It can solve extremely complex problems in a fraction of a second, benefiting portfolio selection, asset allocation, and overall business optimization programs. Other use cases could include ultra-sophisticated underwriting, anti-money-laundering initiatives, anti-fraud neural nets in real time, synthesis of massive amounts of global data, and the development of next-gen encryption and financial cybersecurity technologies.
- **Predictive Analytics-** Fintech companies are embracing predictive analytics, powered by AI, machine learning, and big data, to make informed predictions about future trends, customer behavior, and market dynamics. With the increasing volume of financial transaction data, this investment in data analytics enables businesses to optimize performance, enhance efficiency, and seize opportunities in a rapidly changing market. These analytics are often integrated with big data and data science, using techniques like deep learning, machine learning algorithms, and statistical models such as logistic and linear regression.
- **Blockchain-** Blockchain technology is no longer limited to monetary transactions and cryptocurrencies. Its inherent characteristics, including transparency and immutability, make it an ideal platform for a variety of use cases. The use of blockchain extends to



product tracking in supply chains, provenance of goods, medical record keeping, and other forms of secure data storage and sharing.

- **Cloud-Based Solutions-** Cloud-based solutions dramatically reduce the costs of acquiring, maintaining, and upgrading expensive hardware and software infrastructure. This allows fintech startups and established financial institutions to allocate resources more efficiently. In the highly competitive fintech landscape, cost savings are a crucial factor for innovation and profitability. : Cloud computing's ability to scale IT resources up or down according to demand is invaluable in fintech development. As fintech services expand or experience fluctuations in user activity, the cloud allows for seamless adjustments in computing power and storage. This scalability ensures that fintech solutions remain responsive and cost-effective as they grow. Additionally, Cloud companies invest significantly in cybersecurity measures to protect their clients' data and infrastructure. These measures include encryption, multi-factor authentication, regular backups, and adherence to stringent security standards. By leveraging these cloud-based security features, fintech companies can focus on their core services while maintaining the integrity and confidentiality of financial data.



5B. Computing Hardware

Industry Brief

The computing hardware subsector encompasses all of the physical components that power computing. There are dozens of products that fall under this category, here are some: GPUs, CPUs, Routers, Memory, Servers, and Motherboards. Hardware impacts all mediums of technology from smartphones to desktops to cloud computing.

Key Terms

- Data Center – Large server rooms, usually the size of a football field, that serves as a way to process or store data remotely to companies. The key owners and users of data centers are for cloud computing: Amazon Web Services (AWS), Microsoft Azure, and Google Cloud. Large tech companies such as Apple and Meta use and own personal data centers to self-serve for multiple services.
- Fabrication machine (Fab) – A fabrication machine or commonly referred to as a fab, produces semiconductor chips. These machines are very precise and expensive. With frequent innovations in the space they need to be constantly updated to be able to produce new designs. Capital equipment vendors like: KLA, LAM, ASML, and AM, would supply the inputs to make a fab.
- Semiconductor Manufacturer – A semi manufacturing company owns *fabs* and produces chips. The most known name in the space is Taiwan Semiconductor Manufacturing Company (TSM) in which almost all the names below have exposure to this one advanced foundry.
- Semiconductor Designer – A semi designer makes the designs needed to manufacture the chips. They themselves do not manufacture them as it requires a *fabrication machine (Fab)* which is an expensive investment so the manufacturing is outsourced to foundries. Many of the key players are listed and elaborated on below.

Key Players

Taiwan Semiconductor Manufacturing Company Limited (NYSE: TSM)

As a global semiconductor manufacturing powerhouse, TSMC stands at the forefront of technological innovation. Renowned for its cutting-edge semiconductor *manufacturing*, TSMC has consistently delivered high-performance chips, contributing to its robust market presence. The company's stock performance reflects its resilience and growth, driven by the increasing demand for advanced chip technologies globally. Companies that rely on TSMC for semiconductor manufacturing include NVidia, Apple, AMD, and Intel. TSMC's strategic importance in the semiconductor supply chain underscores its influence and relevance in



driving progress within the industry.

NVIDIA (NASDAQ: NVDA)

NVIDIA is a semiconductor *designer* known for its high end performance GPUs. These GPUs have been the best for 24 years and with the recent push towards AI, it has grown ~117% in the past 2 years and 274% in the past year. Also introducing TPUs in 2020, the A100 chip serves as a rival for Google's current leading TPU v4. NVDA is constantly innovating. Their other product segments include: data center computing, EV computing, cryptocurrency mining processors, Jetson embedded systems, and more.

Intel (NASDAQ: INTC)

Intel is a semiconductor *design* and *manufacturer* company known for arguably the most important part of any computer, the CPU. Its CPUs have been top performing for a long time similar to NVDA with GPUs. But since 2017 that has been put into question as AMD has caught up in performance with better price points. With an acquisition of Mobileye in 2018, INTC has entered the autonomous driving market. Even more recently, INTC has invested ~\$30B into their own *fab* to start manufacturing semi chips. Aside from recent news Intel also operates in these segments: data center and AI, networks and edge, accelerated computing and graphics, and their client computing group.

Advanced Micro Devices (NASDAQ: AMD)

Advanced Micro Devices is a semiconductor *designer* known for being second place. But being in second place for both GPUs and CPUs they sell to both markets instead of being a leader in one. However it is now split for CPUs that AMD has caught up in performance with INTC, giving them half of the market share in the past few years compared to 32% in 2019.

Broadcom (NASDAQ: AVGO)

Broadcom is a leader in communication and networking chips for which they *design*. They offer many different enterprise software solutions for: cybersecurity, data center automation, networking, and much more. Currently this is the second best performing company to capture value from the AI hype. From what the team thinks this value has been driven by their exposure to some of the largest cloud computing companies: Amazon's AWS, MSFT Azure, and GOOGL Cloud.

Dell (NYSE: DELL)

Dell Technologies operates across a spectrum of product segments including: personal computers, laptops, servers, networking equipment, storage solutions, and cybersecurity



International Business Machines (NYSE: IBM)

International Business Machines is a blue chip company known for its hardware and software products. Providing computers, servers, and networking equipment, consulting, cloud computing, and business analytics. They also provide business technology and financial service consulting. A stalwart of the industry as they have provided computing hardware for decades.

AI Impact

The influence of AI has been significant within the semiconductor realm, where major designers and manufacturers have experienced substantial growth in their stock prices. This surge is attributed to the sudden spike in demand for AI and AI-related technologies, with semiconductors playing a pivotal role in supporting this industry. In fact, they currently serve as the backbone of the AI sector, given that without adequate computing power, specific AI functions and goals cannot be achieved. Consequently, semiconductors have become indispensable for the industry's future. Notably, key semiconductor chip designers such as NVDA, INTC, AMD, AVGO, and QCOM have capitalized on this demand, capturing significant value. This trend underscores the widespread belief that these companies will emerge as leading providers of chips and software to AI-focused firms.

Trailing Performance of Key Players





5C. Consumer Technology

Industry Brief

The consumer technology industry revolves around the design, development, and sale of technology products and services aimed at individual consumers. Its core products include smartphones, laptops, tablets, smart home devices, wearables, and various software applications - the consumer tech industry commonly refers to products as "smart" if they are technologically advanced versions of something we previously had, e.g. phones, watches, doorbells.

The industry is characterized by rapid innovation, with companies constantly striving to create more advanced, user-friendly, and affordable products that provide entertainment, convenience, and more. Customers in this industry are diverse, ranging from tech enthusiasts and early adopters to everyday consumers seeking gadgets to enhance their personal and professional lives. Consumer technology, with its constant evolution, is shaping the way people interact with technology in their daily lives.

Key Players

Apple (laptops, phones, watches), Samsung (appliances), Philips (lights) Google (laptops, phones, owns Nest: doorbells, cameras, thermostats), Keurig (coffee makers), Microsoft (computers, tablets), Meta (Metaverse products). All of these products in parentheses are considered "smart," and many of them have the word smart in their name.

AI Impact

AI leads to more capabilities for personalized experiences, for example, better recommendations on streaming platforms, more accurate location and traffic data on mapping apps, more capable language bots on language apps like Duolingo, and more advanced offerings with voice assistants (e.g. Siri, smart lights, appliances, etc). Many of these features are already utilizing AI to offer extremely strong and advanced products, but they will only get better with the evolution of AI.

Apple is the biggest player in the consumer tech industry and has created hardware products that are now a necessary part of life. They have an extremely expansive ecosystem of hardware (phones, laptops, watches) and software (music, health, carplay), and with this comes a massive amount of data on their customers. Machine learning (ML) is an integral part of iOS, meaning they are constantly extracting data and learning about users. This allows for high personalization and deeper insights on apps like Health, Maps, and Music. For example, the Health app tells us trends about our exercise patterns for different time ranges. Apple is already using AI and ML to tailor their products to customers, but as it continues to evolve, Apple will be



at the forefront of AI transformation thanks to its extensive data on customers. A quick example of how they will capitalize on AI and their data is their announced plan to offer a coaching service within the Health app. Apple has an industry-leading understanding of their users, and they will continue to learn even more. As their understanding continues to improve and offerings become even more customer-specific, it becomes increasingly expensive for users to move away from the Apple ecosystem. This provides Apple with the opportunity to upsell as they continue to innovate and keep customers coming back.

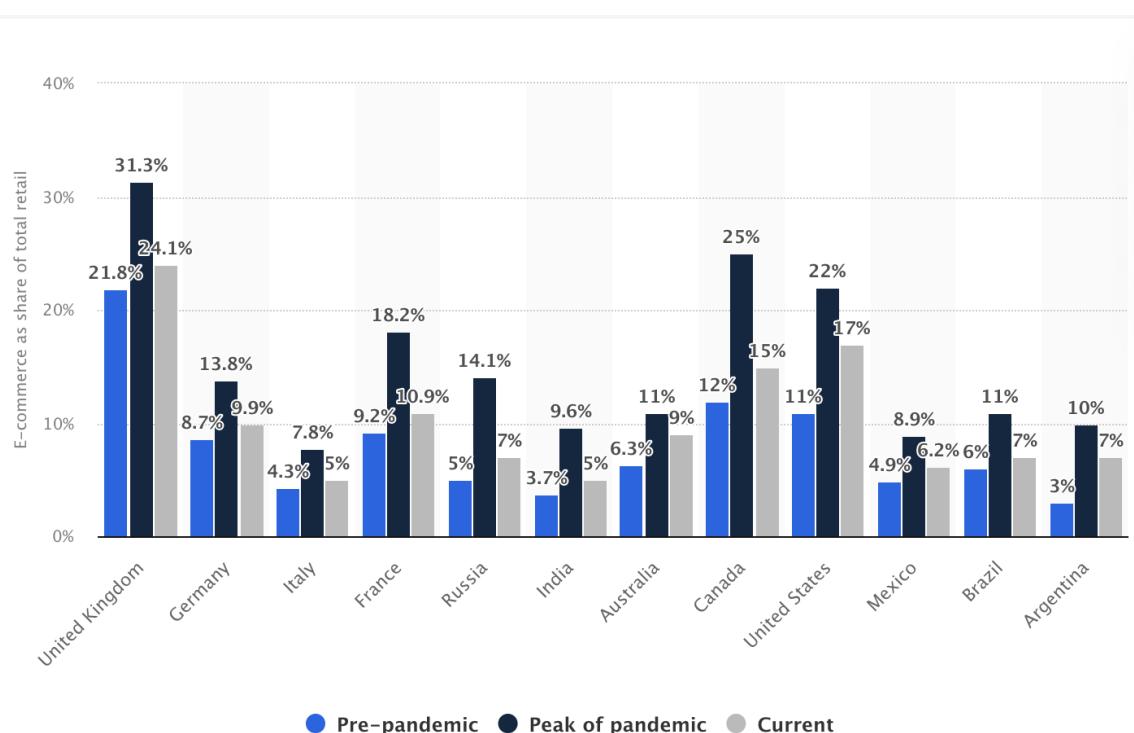
Meta is another giant in the space with a similarly large ecosystem and set of consumer data as Apple. Meta owns Facebook, Instagram, WhatsApp, and more, giving them massive insights into their consumers. They have been using data from these software offerings for years now, especially in the ad space. Apps like Instagram and Facebook tell Meta so much about their customers that they are (and have been) able to have the highest-converting ads of all time. Advertisers have been willing to pay top dollar for access to ad space on these apps. Meta will be able to extract even better insights on their users as AI gets better, which can be used for other products Meta develops and ultimately monetizes. Meta has also begun expanding into the hardware space with their development of VR goggles and headsets. Possibilities for AI applications in the Metaverse seem plentiful, for example, AI avatars that are trained like a language model, meaning they can communicate with users. Meta will also be able to expand their offerings and increase their products' stickiness as their entire ecosystem becomes more tailored to its users.



5D. e-Commerce

Industry Brief

By definition e-Commerce is the buying/selling of goods/services online using one's own website, or established ones like Amazon, Ebay and Shopify. In addition, one can even sell on social media platforms such as Instagram or Facebook Marketplace. There are a number of benefits to be considered when taking a look at the e-Commerce industry. One of the biggest benefits of e-Commerce is the Global Market reach which allows sellers to reach customers from anywhere in the world furthermore expanding the customer base and providing tons of opportunities for growth. This growth was made evident during the pandemic; according to Forbes, U.S. online spending hit record levels and e-Commerce jumped 55% to hit \$1.7 trillion. The number of online shoppers grew by roughly 1 billion from 2019-2022 and it will only continue to grow in the future.



Along with the Global Market reach, e-Commerce allows a faster buying process, easy product and price comparisons and provides a lot of flexibility for customers due to the several different payment methods that are made available to them.



Key Players

Some of the key players in e-Commerce are Amazon and Alibaba. Known throughout the world, Amazon is able to provide a wide range of goods and services to both individuals and companies whether it be through subscription services like Amazon Prime Video or Amazon Prime. In the e-Commerce industry, they currently command a 37.8% market share, positioning them as leaders in retail e-Commerce. Alibaba is the leading e-Commerce and technology company in China and it often gets compared with Amazon because it has a similar business model although it has a wider range of products/services than Amazon. In China Alibaba makes up around 58.2% of market shares. In Q3 FY 2022, the company's revenue surged by 9.7% year over year, showcasing its operational prowess. However, a concerning 74.3% year over year dip in net income, due to escalating costs. On the competitive front, even with its stronghold in China, Alibaba isn't without rivals. Renowned entities like Tencent and Amazon present stiff competition in various sectors. Further shaping its corporate narrative, Alibaba recently underwent a significant leadership transition. Toby Xu is poised to assume the CFO role in April 2022, succeeding Maggie Wu. Such strategic shifts, often seen in corporate giants, are designed to infuse fresh perspectives and propel forward momentum.

AI Impact

Artificial intelligence (AI) has become indispensable in the current digital era, reshaping industries and streamlining operations. The two giants, Alibaba and Amazon, are making tremendous gains in the field of artificial intelligence. Qwen-VL and Gwen-VL-Chat, two ground-breaking AI models, were recently introduced by Alibaba. According to its claims, these models are a step up from traditional text-based interactions since they can understand visuals and facilitate complex conversations:

- Qwen-VL-Chat can decipher a Chinese hospital sign and deliver pertinent information about where the various departments are located.
- Sectors including healthcare, retail, and others that depend on visual information could be completely transformed by such capabilities.
- By empowering researchers worldwide, Alibaba's choice to open-source these models encourages AI innovation without the constraints of time and cost.

Amazon, not lagging in the AI race, announced its investment in the AI startup, Anthropic. This collaboration aims to bolster Amazon's AI capabilities to stand tall against competitors like Microsoft and Google. The investment, amounting to \$1.25 billion initially, can escalate up to \$4 billion, emphasizing Amazon's commitment to AI. Anthropic, known for its generative AI models, aligns perfectly with Amazon's vision to provide diverse AI models to its



customers. For instance, Anthropic's Claude 2, an AI model skilled in analyzing extensive business or legal documents, can revolutionize sectors reliant on comprehensive data analysis. This move may become similar to Microsoft's very own deal with OpenAI as Amazon has significant computing power through their AWS services.



5E. Software as a Service (SaaS)

Industry Brief

SaaS, or Software as a Service, refers to companies whose primary value and revenue source lies in the software they provide. Think of them as intermediaries between software users and producers. Much like Netflix's software connecting content producers with viewers, offering their software as a means to access movies or services. Users subscribe to these platforms to utilize their software, paying a regular fee for access.

In this context, we are focusing on B2B SaaS, which caters to businesses by offering various software services. Companies pay subscription fees for these services, much like consumers paying for Netflix. For instance, Shopify provides foundational e-commerce software for businesses to build websites, Zoom specializes in video conferencing software, and Salesforce offers customer relationship management software for sales teams. These companies, often dubbed "darlings" in the eyes of investors, generate substantial revenue once their software gains a substantial user base and outperforms competitors.



To maintain their position and profitability, SaaS companies must continuously update and enhance their software, much like Netflix's regular addition of features to retain subscribers. Achieving this requires a skilled team of engineers and sales representatives. Engineers ensure the platform's operational excellence, while sales teams scout for new clients. A top-tier product often attracts customers due to its reputation. The rise of AI is another essential factor in the SaaS industry, as many startups offer specialized "AI for x" services, such as DataSnipper's "AI for audit," attracting subscription fees and investor capital.



Industry Insights

The Cloud-Based Advantage

SaaS companies predominantly operate in the cloud, enabling users worldwide to access their software with an internet connection. This innovation has replaced the need to download software or install physical hardware. The cloud facilitates automatic updates and greater efficiency. Businesses are increasingly moving their operations to the cloud, creating significant opportunities for SaaS providers. For instance, CrowdStrike offers cybersecurity for various endpoints, while DocuSign simplifies document signing worldwide.

The Role of Mission Critical Software

The success of SaaS companies often hinges on the mission-critical nature of their products. Tools like Salesforce significantly ease operations for sales teams, making them willing to pay substantial subscription fees. Although alternative methods like Microsoft Excel or physical documents are available, the user-friendliness of SaaS solutions drives their adoption.

Analyzing the SaaS Industry

To assess a SaaS company's potential, we must examine IT budgets, as these teams are the ones subscribing to these services. A higher budget typically correlates with increased spending on SaaS solutions. In specific verticals, identifying the leading players is crucial. For instance, Salesforce dominates the CRM market, while the video conferencing market is more saturated, featuring competitors like Apple FaceTime, Google Meet, and Skype. In this crowded space, Zoom stands out as the B2B leader, focusing exclusively on professional conferences and offering tools like screenshare.

Key Metrics

Annual Recurring Revenue (ARR) is a vital metric in the SaaS industry. It represents how much revenue a company generates annually, which is particularly valuable for subscription-based businesses. Net Retention, on the other hand, measures how effectively a company retains its customers. When this number exceeds 100%, it indicates that the company not only retains all its revenue but also adds more through additional revenue, signaling healthy growth.

Morgan Stanley CIO Survey Results

Morgan Stanley surveyed Chief Information Officers (CIOs) from some of the largest companies in the world. CIOs largely determine investment into information technology (IT) which largely comprises of SaaS spending.



IT Budget Trends

IT budgets are expected to grow by 3.4% next year, indicating a positive shift. Historically, these budgets have grown at an average rate of approximately 4%, which is a notable increase from the 2% growth experienced last year and the 2.7% growth in 2023. However, it's worth noting that this current growth rate is still below the long-term historical average.

Investment in AI and LLM

A substantial portion of this growth in IT budgets is attributed to investments in long-term Artificial Intelligence (AI) and Legal Lifecycle Management (LLM) spending. These emerging technologies and solutions are driving the budgetary priorities for many organizations as they plan for the future.

Macroeconomic Pressures and Company Performance

One significant aspect of the survey analysis pertains to macro pressures. Although these broader economic challenges aren't necessarily improving, companies are displaying a remarkable ability to outperform expectations. This suggests that macro pressures may not be as dire as initially anticipated.



6. Sources

What Are Neural Networks?. What Are Neural Networks?. 21 Dec. 2023,
[https://www.ibm.com/topics/neural-networks.](https://www.ibm.com/topics/neural-networks)

How Large Language Models Work. How Large Language Models Work. 21 Dec. 2023,
[https://www.youtube.com/watch?v=5sLYAQS9sWQ.](https://www.youtube.com/watch?v=5sLYAQS9sWQ)

Introduction to Large Language Models. Introduction to Large Language Models. 21 Dec. 2023,
[https://www.youtube.com/watch?v=zizonToFXDs.](https://www.youtube.com/watch?v=zizonToFXDs)

Openai Customer Story: Duolingo. Openai Customer Story: Duolingo. 21 Dec. 2023,
[https://openai.com/customer-stories/duolingo.](https://openai.com/customer-stories/duolingo)

Fine-tuning Large Language Models (llms) | W/ Example Code. Fine-tuning Large Language Models (llms) | W/ Example Code. 21 Dec. 2023,
[https://www.youtube.com/watch?v=eC6Hd1hFvos.](https://www.youtube.com/watch?v=eC6Hd1hFvos)

Visa Inc. “Visa Launches \$100 Million Generative AI Ventures Initiative.” *Usa.visa.com*, 2 Oct. 2023, usa.visa.com/about-visa/newsroom/press-releases.releaseld.19986.html. Accessed 10 Oct. 2023.

Software as a Service [saas] Market Size & Growth, 2030. Software as a Service [saas] Market Size & Growth, 2030. 21 Dec. 2030,
[https://www.fortunebusinessinsights.com/software-as-a-service-saas-market-102222.](https://www.fortunebusinessinsights.com/software-as-a-service-saas-market-102222)

State of SaaS | SaaS Industry Reports | Silicon Valley Bank. State of SaaS | SaaS Industry Reports | Silicon Valley Bank. 21 Dec. 2023,
[https://www.svb.com/trends-insights/reports/state-of-saas-report/.](https://www.svb.com/trends-insights/reports/state-of-saas-report/)

Software as A Service (saas) Market Size Report, 2030. Software as A Service (saas) Market Size Report, 2030. 21 Dec. 2030,
[https://www.grandviewresearch.com/industry-analysis/saas-market-report.](https://www.grandviewresearch.com/industry-analysis/saas-market-report)

Drive Growth by Building New SaaS Businesses | McKinsey. Drive Growth by Building New SaaS Businesses | McKinsey. 21 Dec. 2023,
[https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-saas-factor-six-ways-to-drive-growth-by-building-new-saas-businesses.](https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-saas-factor-six-ways-to-drive-growth-by-building-new-saas-businesses)



Clouded Judgement. *Clouded Judgement.* 21 Dec. 2023,
<https://cloudedjudgement.substack.com/>.

SAP Brandvoice: Making ERP A True System of Intelligence in the Cloud. *SAP Brandvoice: Making ERP A True System of Intelligence in the Cloud.* 21 Dec. 2023,
<https://www.forbes.com/sites/sap/2023/10/02/making-erp-a-true-system-of-intelligence-in-the-cloud/>.

Amazon Vs Walmart: The Battle for the Ecommerce Crown. *Amazon Vs Walmart: The Battle for the Ecommerce Crown.* 21 Dec. 2023,
<https://etaileast.wbresearch.com/blog/walmart-not-surviving-but-thriving-strategy-ecommerce-war-against-amazon>.

Amazing Amazon Statistics You Need to Know to Amplify Growth in 2023. *Amazing Amazon Statistics You Need to Know to Amplify Growth in 2023.* 21 Dec. 2023,
<https://influencermarketinghub.com/amazon-statistics/>.

Best E-commerce Companies in 2023. *Best E-commerce Companies in 2023.* 21 Dec. 2023,
<https://www.fool.com/investing/stock-market/market-sectors/consumer-discretionary/top-e-commerce-companies/>.

Introduction to Cloud TPU. *Introduction to Cloud TPU.* 21 Dec. 2023,
<https://cloud.google.com/tpu/docs/intro-to-tpu>.

What Is Hyperscale and Who Are the Hyperscalers?. *What Is Hyperscale and Who Are the Hyperscalers?.* 21 Dec. 2023, <https://dgtlinfra.com/hyperscale-and-hyperscalers/>.

What Is Hyperscale?. *What Is Hyperscale?.* 21 Dec. 2023,
<https://www.digitalrealty.com/resources/articles/what-is-hyperscale?t=1703193178664?l=atest>.

How to Design and Build a Data Center. *How to Design and Build a Data Center.* 21 Dec. 2023,
<https://www.techtarget.com/searchdatacenter/How-to-design-and-build-a-data-center>.

What Does It Cost to Build a Data Centre?. *What Does It Cost to Build a Data Centre?.* 21 Dec. 2023, <https://siteltd.co.uk/blog/what-does-it-cost-to-build-a-data-centre/>.



Who Is Winning the Microsoft-openai Partnership?. Who Is Winning the Microsoft-openai Partnership?. 21 Dec. 2023,

<https://fortune.com/2023/01/24/whos-getting-the-better-deal-in-microsofts-10-billion-tie-up-with-chatgpt-creator-openai/>.

Who Owns Openai?. Who Owns Openai?. 21 Dec. 2023,

<https://fourweekmba.com/who-owns-openai/>.

Openai Could See Its Secondary-market Valuation Soar to \$90B. Openai Could See Its Secondary-market Valuation Soar to \$90B. 21 Dec. 2023,

<https://techcrunch.com/2023/09/26/openai-is-reportedly-raising-funds-at-a-valuation-of-80-billion-to-90-billion/>.

What Is a Dataset in Machine Learning. What Is a Dataset in Machine Learning. 21 Dec. 2023,

<https://labelyourdata.com/articles/what-is-dataset-in-machine-learning>.

Bloomberg - Are You a Robot?. Bloomberg - Are You a Robot?. 21 Dec. 2023,

<https://www.bloomberg.com/tosv2.html?vid=&uuid=f6430db0-a045-11ee-9734-197b22caeab9&url=L25Id3MvYXJ0aWNsZXMvMjAyMy0wNC0yNS9hcHBsZS1hYXBsLWRldmVsb3BpbmctYWktaGVhbHRoLWNvYWNoaW5nLXNlcnpY2UtaXBhZG9zLTE3LWhlYWx0aC1hcHA/ZW1iZWRkZWQtY2hlY2tvdXQ9dHJ1ZQ==>

Global Search Engine Desktop Market Share 2023. Global Search Engine Desktop Market Share 2023. 21 Dec. 2023,

<https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>.

Artificial Intelligence - Turing Test, Machine Learning, Robotics / Britannica. Artificial Intelligence - Turing Test, Machine Learning, Robotics / Britannica. 21 Dec. 2023,

<https://www.britannica.com/technology/artificial-intelligence/The-Turing-test>.

Microsoft and Meta Expand Their AI Partnership with Llama 2 on Azure and Windows. Microsoft and Meta Expand Their AI Partnership with Llama 2 on Azure and Windows. 21 Dec. 2023,

<https://blogs.microsoft.com/blog/2023/07/18/microsoft-and-meta-expand-their-ai-partnership-with-llama-2-on-azure-and-windows/>.

Microsoft's \$13 Billion Bet on Openai Carries Huge Potential Along with Plenty of Uncertainty.

21 Dec. 2023,



<https://www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html>.

Oremus, Will. "Analysis | Google's AI Passed a Famous Test — and Showed How the Test Is Broken." *Washington Post*, 17 June 2022, www.washingtonpost.com/technology/2022/06/17/google-ai-lamda-turing-test/.

Goertzel, Ben. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*. 10.2478/jagi-2014-0001.

Patel, Ankur A. "LLaMA: Meta's Open-Source Rival to Google and OpenAI." www.ankursnewsletter.com/p/llama-metas-open-source-rival-to#:~:text=LLaMA%20is%20more%20also%20outperforms. Accessed 12 Oct. 2023.

Southern, Matt G. "Google Working to Remove Bard Chat Transcripts from Search." *Search Engine Journal*, 29 Sept. 2023, www.searchenginejournal.com/google-working-to-remove-bard-chat-transcripts-from-search/497444/#close. Accessed 30 Sept. 2023.

Tiku, Nitasha. "The Google Engineer Who Thinks the Company's AI Has Come to Life." *Washington Post*, The Washington Post, 11 June 2022, www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

Weitzman, Tyler. "Council Post: GPT-4 Released: What It Means for the Future of Your Business." *Forbes*, 28 Mar. 2023, www.forbes.com/sites/forbesbusinesscouncil/2023/03/28/gpt-4-released-what-it-means-for-the-future-of-your-business/?sh=406e82e2dc6a. Accessed 12 Oct. 2023.

Goyal, Deepak, et al. *Reimagining the Future of Finance*. 2023.

Kunduru, A. R. "ARTIFICIAL INTELLIGENCE ADVANTAGES IN CLOUD FINTECH APPLICATION SECURITY". *CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES*, Vol. 4, no. 8, Aug. 2023, pp. 48-53, <https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/492>.

Mastercard Inc. "Mastercard Leverages Its AI Capabilities to Fight Real-Time Payment Scams." www.mastercard.com/news/press/2023/july/mastercard-leverages-its-ai-capabilities-to-fight-real-time-payment-scams.



Miquid, and Dominika Słomka. "Predictive Analytics in Fintech: Benefits, Use Cases | Miquid Blog." *Miquid*, 1 Aug. 2023, www.miquid.com/blog/predictive-analytics-in-fintech/.

Taherdoost, Hamed, Blockchain: A Catalyst in Fintech Future Revolution (2022). Future Technology (FUTECH) 2023, Available at SSRN: <https://ssrn.com/abstract=428586>.