Artificial Intelligence in 2023:

An Overview of the Current State of Al, Investment Trends and Recommendations for AVP

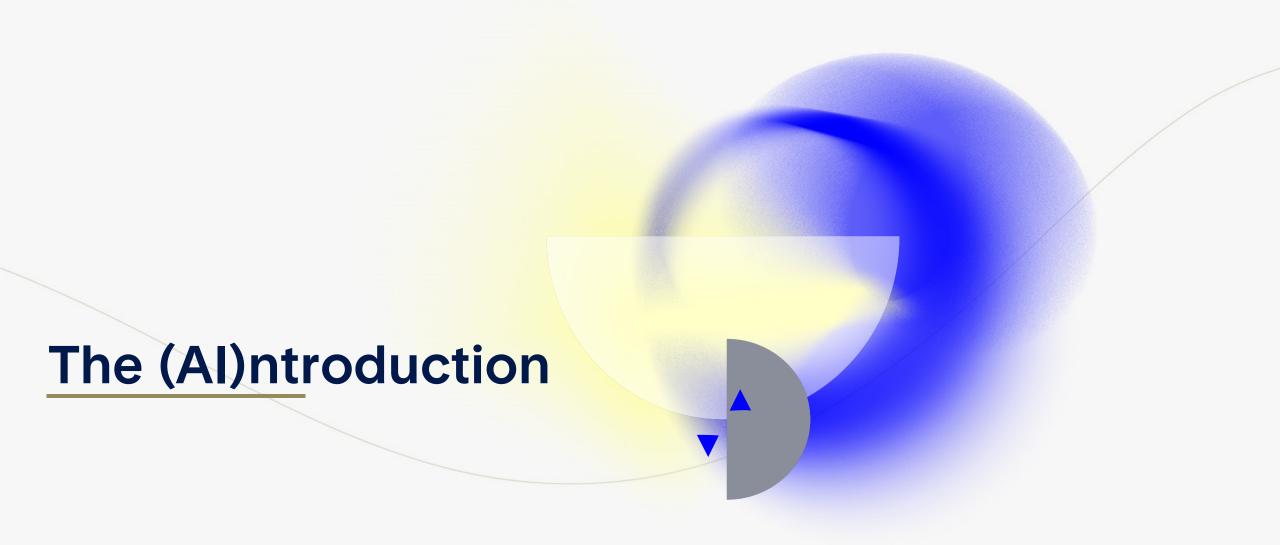
We invest in great entrepreneurs
We support outstanding companies



Table of Contents

Presentation Section

	The (AI)ntroduction	3
	Executive Summary	7
	Artificial Intelligence – Why Now?	9
IV	The Market for Al	14
V	Al Investment Trends & Metrics	17
VI	AVP's Al Market Map & Diligence Checklist	22
VII	ESG Considerations	26
VIII	Final Thoughts & Next Steps	30
IX	Appendix	
	Al Architecture & Techniques	34
	Overview of Generative Al	42
	Multiples Sensitivity by Year	49



A Welcome Message



What do these people have in common?



















It's AVP's New York Al-Generated Team!





















Executive Summary

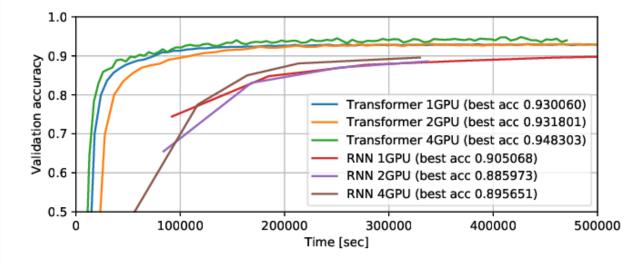
- Over the past five years, artificial intelligence (AI) has seen a step-function acceleration in technological progress due to three key factors
 - Advancements in ML architecture (e.g., Transformers)
 - Compute power has increased exponentially
 - Increased creation of and access to data (for model training)
- Wall Street Research analysts are bullish on the future of Al reports suggest that Al is currently a >\$30 billion market with 100% expected annual growth
 - Widespread Al adoption could eventually drive a 7%, or almost \$7 trillion, increase in annual global GDP over a 10-year period
- Venture capital investment in Al has boomed since 2019
 - Between 2019 and 2022, there have been 14,000+ unique VC funding rounds globally for AI companies (~5,500 of these rounds occurred in either the United States, Canada, Europe or Israel)^(a)
 - Over the same period, both the funding amount and valuation of rounds have steadily increased in conjunction with the broader tech bull market landscape
- For AVP, there is ample opportunity to invest in AI within the application layer (i.e., SaaS applications that use AI in their core product offering)
 - However, AVP should also opportunistically consider companies in the foundation model and AI operations layers as companies in these segments will provide the core infrastructure and tooling on which the application layer is built
- Al is still very much in the early innings and new advancements are happening at a faster pace we view Al as the next
 generational technology shift and should seek to invest in the category in the near term

Artificial Intelligence – Why Now?

Why is Al Interesting Now?



- In 2017, Google introduced the Transformer model, a novel neural network architecture that requires less computation to train and is better suited for modern machine learning hardware
 - The key innovation in the Transformer model revolved around the "self-attention mechanism", which allows Transformer models to understand relationships between words at greater distances
 - Moreover, Transformers can be parallelized across both training and inference, significantly speeding up training times
- The Transformer model improved upon existing neural network architectures (CNNs and RNNs) by addressing limitations that each model faces
 - Convolutional Neural Networks (CNNs) focus on relative location and proximity (vs. an entire sequence of data)
 - Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTMs), have memory limitations and tend to read in one direction
- The advent of transformers has led to the proliferation of large language models (LLMs), including OpenAl's GPT-4 and Google's LaMDA

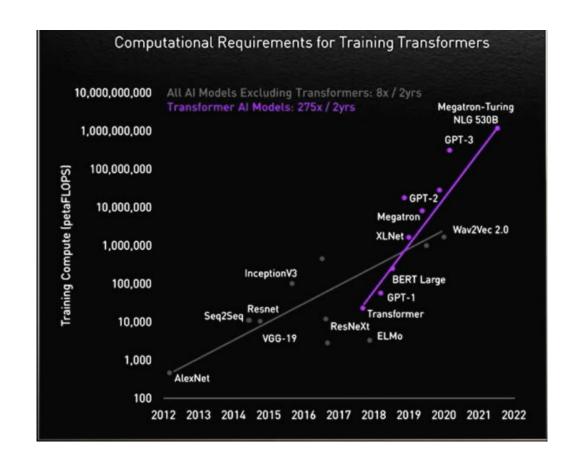


Source: Google Research and other publicly available information.

Why is AI Interesting Now? (cont'd)

2 Compute power has increased exponentially

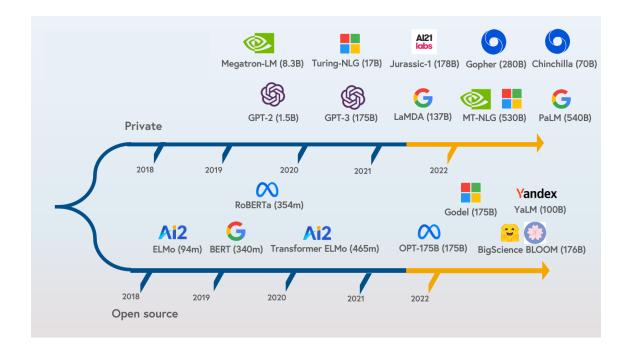
- Since 2012, the amount of compute used in the largest Al training runs has been increasing exponentially with a 3.4-month doubling time (by comparison, Moore's Law had a 2-year doubling period)
 - This metric has grown by more than 300,000x (a 2-year doubling period would yield only a 7x increase)
 - Within many current domains, more compute seems to lead predictably to better performance, and is often complementary to algorithmic advances
- The increase in compute has been largely driven by two factors
 - Researchers repeatedly finding ways to use more chips in parallel and being willing to pay the economic cost of doing so
 - Custom hardware that allows more operations to be performed per second for a given price (GPUs and TPUs, or Tensor Processing Units)
- As compute power has increased, so too have the parameters of large language models
 - For example, GPT-3 was trained on 175 billion parameters (vs. 1.5 billion for GPT-2)



Source: OpenAI and other publicly available information.

Why is Al Interesting Now? (cont'd)

- 3 Increased creation of and access to data (for model training)
- Since the advent of the Transformer model in 2017, several companies including OpenAl, Google, Microsoft, Meta and Nvidia have invested resources towards building LLMs
 - The cost of training these models is incredibly expensive and requires extensive compute power – for the average company, access to both capital and compute is limited
 - For context, the cost of a single Nvidia GPU server can cost upwards of \$200,000 with each server consuming up to 6.5 kilowatts
- Developers of LLMs have taken the route of both open source software and private APIs to democratize access to models and data for companies looking to build on top of their infrastructure
 - Open Source: Hugging Face offers over 120k models, 20k datasets and 50k demos which engineers can easily use to begin their software development
 - Private: Open AI offers its LLM capabilities via an API, which can cost anywhere from \$0.0004 to \$0.02 per API call



12

What are market leaders saying?



The development of AI is as <u>fundamental as the creation of the microprocessor, the personal computer, the Internet, and the mobile phone</u>. It will change the way people work, learn, travel, get health care, and communicate with each other. Entire industries will reorient around it. Businesses will distinguish themselves by how well they use it. There will be an explosion of companies working on new uses of AI as well as ways to improve the technology itself. But one big open question is whether we'll need many of these specialized AIs for different uses or whether it will be possible to develop an artificial general intelligence that can learn any task.



With regard to AI, it is a major focus of ours. It's incredible in terms of how it can enrich customers' lives. And you can look no further than some of the things that we announced in the fall with crash detection and fall detection or back a ways with ECG. I mean these things have literally save people's lives. And so we see an enormous potential in this space to affect virtually everything we do. It's obviously a horizontal technology, not a vertical. And so it will affect every product in every service that we have.



The age of Al is upon us and Microsoft is powering it. We are witnessing nonlinear improvements in capability of foundation models, which we are making available as platforms. **We fully expect us to sort of incorporate Al in every layer of the stack**, whether it's in productivity, whether it's in our consumer services. For the last 3.5, 4 years, we've been working very, very hard to build both the training supercomputers and now, of course, the inference infrastructure because once you use Al inside of your applications, it goes from just being training-heavy to inference.



First, the Al opportunity ahead. Al *is the most profound technology we are working on today*. Our talented researchers, infrastructure and technology make us extremely well positioned as Al reaches an inflection point. Already, breakthroughs in everything from natural language understanding to generative Al are fueling our ability to deliver results that drive meaningful performance for advertisers and are useful to users. In fact, our Transformer research project and our field-defining paper in 2017 as well as our path-breaking work in diffusion models are now the basis of many of the generative Al applications you're starting to see today.

Source: Public filings and other publicly available information.



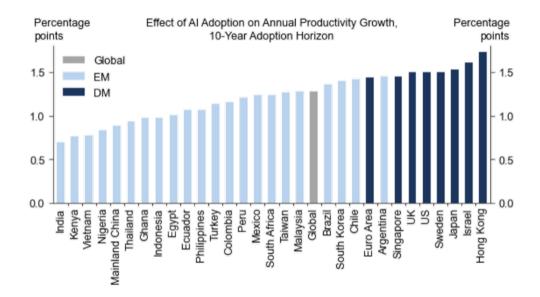
Wall Street is Bullish on the Future of Al

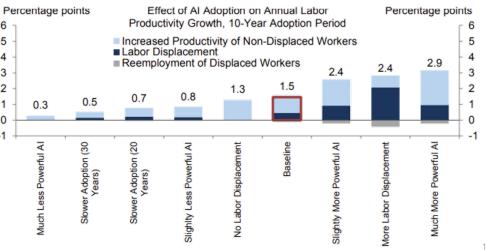
Oppenheimer Equity Research (March 9, 2023)

- We estimate that AI will drive **half the incremental GDP** over the next decade, **representing 20% of global GDP in 2032**
- The current Al market is only in the \$30bn range, set to grow 100% per year for a while, but the impact on the economy will be 10x this
- ChatGPT is an order of magnitude better than anything that has come before it and driving the reengineering of cloud infrastructure
- Over the next decade, most of the value in AI will accrue to platforms that have the **capacity to support the training compute demand**

Goldman Sachs Equity Research (March 26, 2023)

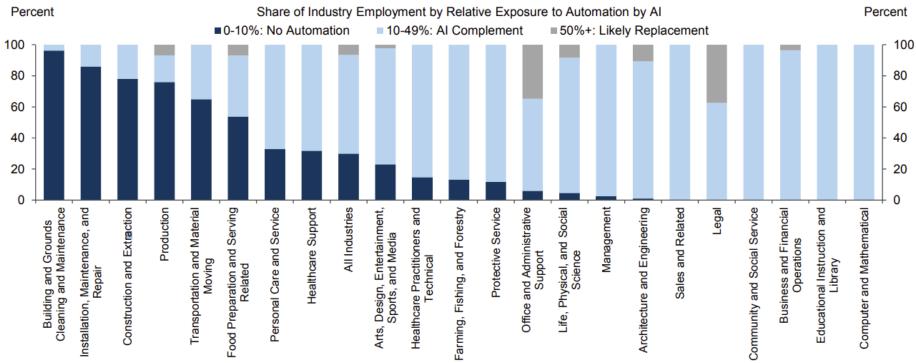
- Widespread Al adoption could eventually drive a 7%, or almost \$7
 trillion, increase in annual global GDP over a 10-year period
- Global private investment in AI represented \$94 billion in 2021; if investment in AI continues to increase at the more modest pace that software investment grew during the 1990s, US investment in AI alone could approach 1% of US GDP by 2030
- We estimate that generative Al could boost aggregate labor productivity by 1.5% in the US; however, more powerful Al could drive productivity gains to 3%





Al as a Complement, Not a Substitute

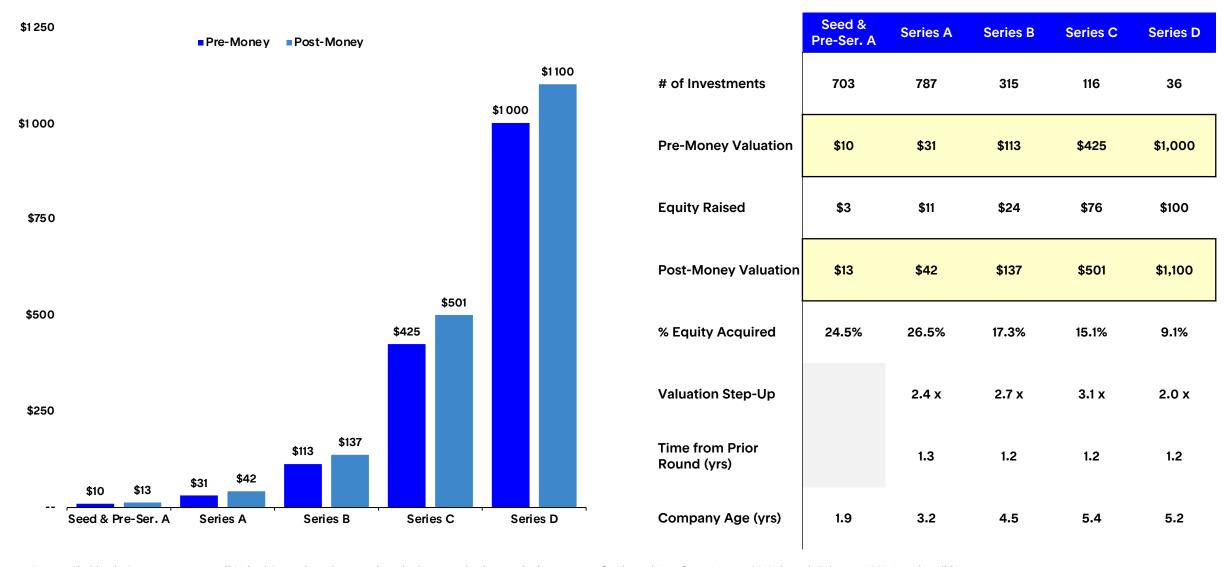
- In the US and Europe, roughly 67% of current jobs are exposed to some degree of Al automation and generative Al could substitute up to 25% of current work
- Although the impact of Al on the labor market is likely to be significant, most jobs and industries are only partially exposed to automation and are thus more likely to be complemented rather than substituted by Al
 - 7% of current US employment will be substituted by AI, 63% complemented and 30% unaffected
 - Replacement in legal and administrative fields, little effect in manual and outdoor jobs and productivity-enhancement everywhere else
- The combination of labor cost savings, new job creation and higher productivity for non-displaced workers raises the possibility of a
 productivity become that raises economic growth substantially.



Source: Goldman Sachs Global Investment Research.

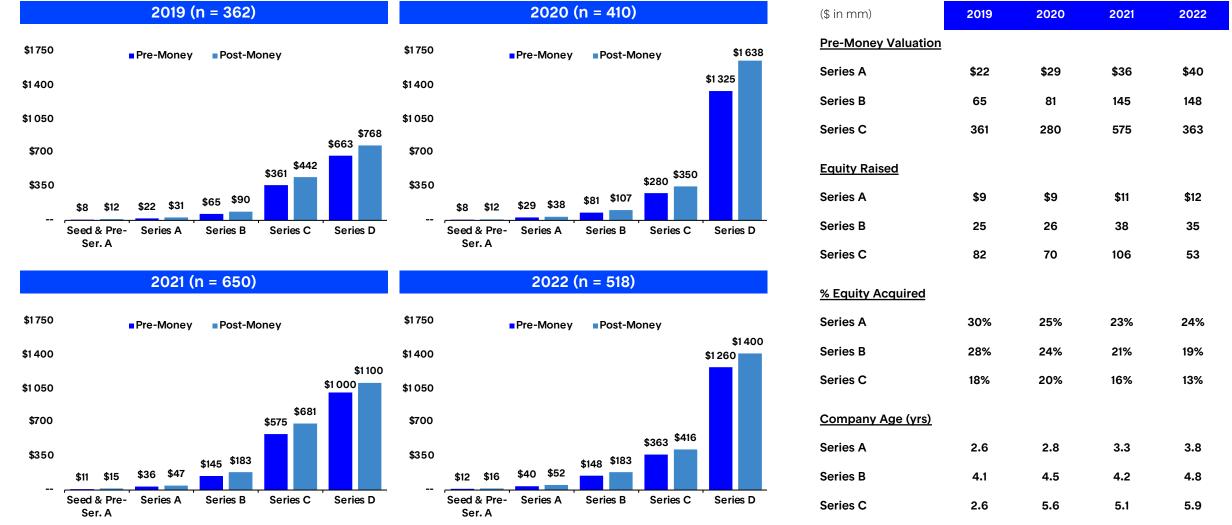
Al Investment Trends & Metrics

Trends and Stats in Artificial Intelligence Investing (Jan. 2019 – Feb. 2023)



Trends and Stats in Artificial Intelligence Investing (Jan. 2019 – Feb. 2023) (cont'd)

Series A and B median entry valuations have doubled from 2019 levels; Series A and Series B median round sizes have increased by 33% and 40% since 2019, respectively



Source: Pitchbook. Screen represents all United States, artificial intelligence companies who have received an equity investment from January 2019 through February 2023. Incudes all investments received by a company during the period, not just the most recent investment. Does not include companies who did not disclose the amount raised. Figures represent medians.

Illustrative Al Investment Sensitivity – Entry Revenue Multiples (Jan. 2019 – Feb. 2023)

			Series	A		
(\$ in mm)				ARR		
		\$1.0	\$1.5	\$2.0	\$2.5	\$3.0
Pre-Money Valuation	\$25	25 x	17 x	12 x	10 x	8 x
	\$28	28 x	19 x	14 x	11 x	9 x
	\$31	31 x	21 x	16 x	12 x	10 x
re-	\$34	34 x	23 x	17 x	14 x	11 x
Δ _	\$37	37 x	25 x	19 x	15 x	12 x

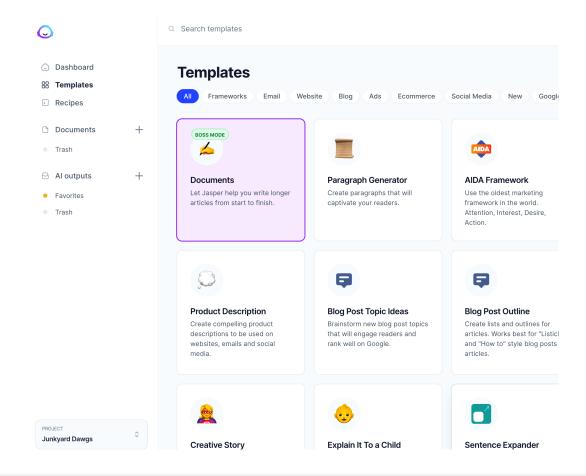
Series B							
(\$ in mm)		ARR					
		\$2.0	\$4.0	\$6.0	\$8.0	\$10.0	
Pre-Money Valuation	\$91	45 x	23 x	15 x	11 x	9 x	
	\$102	51 x	26 x	17 x	13 x	10 x	
	\$113	57 x	28 x	19 x	14 x	11 x	
-all	\$125	62 x	31 x	21 x	16 x	12 x	
a –	\$136	68 x	34 x	23 x	17 x	14 x	

Series C							
(\$ in mm)		ARR					
		\$8.0	\$11.0	\$14.0	\$17.0	\$20.0	
Pre-Money Valuation	\$340	42 x	31 x	24 x	20 x	17 x	
	\$382	48 x	35 x	27 x	22 x	19 x	
	\$425	53 x	39 x	30 x	25 x	21 x	
	\$467	58 x	42 x	33 x	27 x	23 x	
	\$510	64 x	46 x	36 x	30 x	25 x	

Series D							
(\$ in mm)		ARR					
		\$20.0	\$25.0	\$30.0	\$35.0	\$40.0	
Pre-Money Valuation	\$800	40 x	32 x	27 x	23 x	20 x	
	\$900	45 x	36 x	30 x	26 x	23 x	
	\$1,000	50 x	40 x	33 x	29 x	25 x	
	\$1,100	55 x	44 x	37 x	31 x	28 x	
Δ -	\$1,200	60 x	48 x	40 x	34 x	30 x	

Recent Al Investment Case Study: Jasper

- In October 2022, Jasper raised a \$125 million Series A at a \$1.5 billion post-money valuation
 - The investment was led by Insight Partners with participation from Bessemer Venture Partners, IVP, Coatue and others
- Jasper leverages AI to generate content for blog articles, social media posts and website copy
 - Using the platform, customers can describe in natural language what they want Jasper to write, whether a keyword-rich piece designed to rank well in search engines or existing content repurposed with additional context
 - Jasper's language models are trained on 10% of the web and are fine-tuned for "customer specificity"
- According to the company, Jasper has 70,000+ customers and generated \$45 million in revenue in 2021 and expects to reach \$75 - \$90 million in revenue in 2022
 - Implied Pre-Money Multiple (2021): 31x
 - Implied Pre-Money Multiple (2022): 15x 18x



"With the advent of OpenAl's GPT-3, we saw an opportunity to launch an Al content platform that could help businesses and professional creators brainstorm and do their work more quickly and efficiently. The folks that will win at generative Al will be the ones that have the best feedback loops – we're committed to building the best feedback to Al loop." – Dave Rogenmoser, CEO

AVP's Al Market Map & Diligence Checklist

Basis of Presentation

- The following page presents AVP's Al Market Map^(a), which is segmented into four categories
 - Cross-Industry Applications
 - Verticalized Solutions
 - Generative Al
 - Developer Tools, Model Deployment and Operations
- The market map was largely constructed based on detailed Pitchbook screening and supplemented by other publicly available Al market maps / industry reports
 - Pitchbook Screen Criteria:
 - Pitchbook Industry Tag / Keywords: "Al", "Artificial Intelligence", "Deep Learning"
 - Year Founded: 2012 2022
 - Geography (HQ): United States, Canada, Europe, Israel
 - Financing Round: Seed, Series A, Series B, Series C
 - Deal Size(b): \$2mm \$60mm
 - Total Capital Raised: \$3mm \$75mm
 - Post-Money Valuation: \$5mm \$400mm
 - Employee Count: 10 250
 - Investor-Backing: Venture Capital
 - The market map is specific to AVP in the sense that it generally focuses on companies that may be interesting for AVP's early stage and growth funds^(c)

⁽a) Market map was made by AVP and is not a duplicate of an industry report / banker presentation. While third party reports were referenced as part of the creation of the market map, all research was conducted by AVP employees.

⁽b) Financing size for any round a company received, not just the latest funding round.

⁽c) For Generative AI specifically, some companies may be earlier stage than AVP typically looks at given the industry is very nascent.

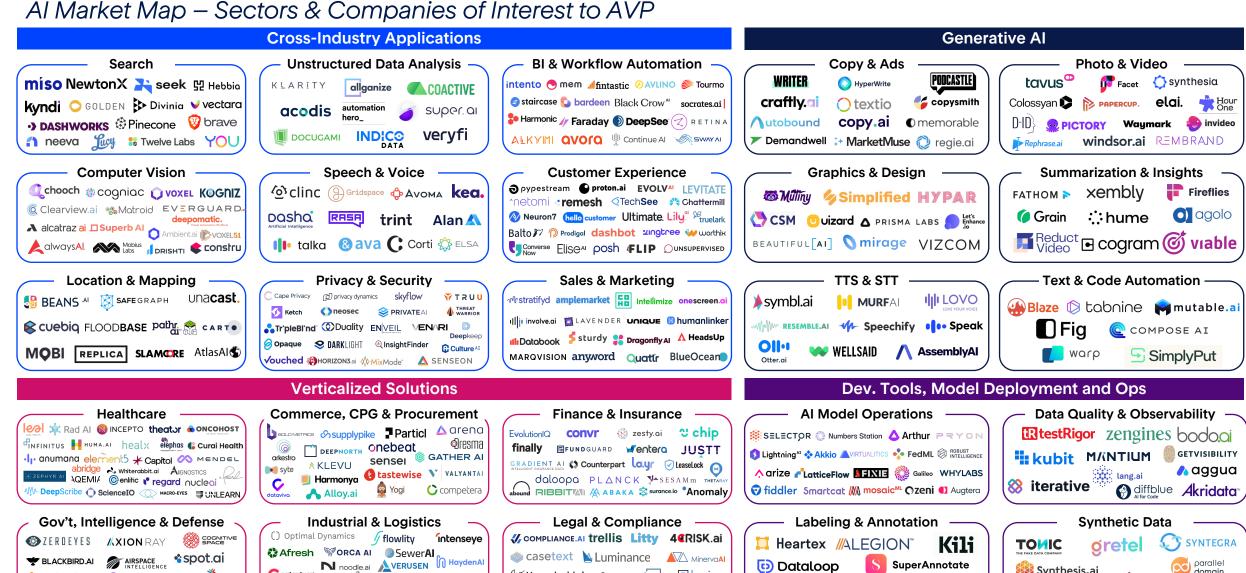
🔌 actuate

MODERN

Sentinel Sencity NOTRAFFIC percipient.ai

greyparrot EXPEDOCK NONEL

SWAPP ALICE ** Zendrive



REGOLOGY PACTUM On JUIO laurel

(ii) Dataloop

@NCORD

SuperAnnotate

🗲 surge" 🦰 datasaur.ai

Synthesis.ai

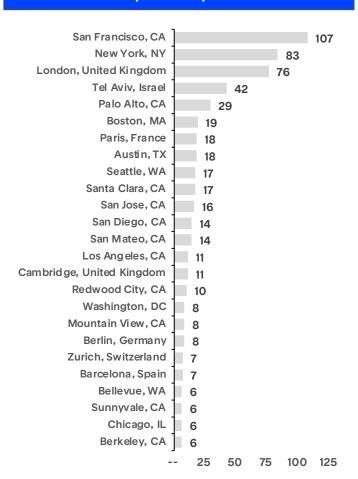
Diveplane

datagen

Al Market Map (cont'd)

Nearly 1,000 artificial intelligence companies generally fall within AVP's early stage and growth stage criteria

Al Companies by Location





Select AVP Experience with AI Companies

Sector: Generative AI - Photo & Video

Geography: Israel

Company Description: D-ID's creative AI technology takes images of faces and turns them into high-quality, photorealistic videos.

D-ID's technology can combine images with audio or text to give them expression and speech.

AVP Experience: US Early Stage team led D-ID's Series A in 2020 (follow-on investment in Series B in Q1 2022)



Sector: Verticalized Solutions - Healthcare

Geography: France

Company Description: Incepto offers artificial intelligence-based medical imaging applications designed to offer computed

tomography scans, chest x-ray and detections for conditions such as fracture and bowel occlusion.

AVP Experience: EU Early Stage team led Incepto's Series A in 2019 (follow-on investment in Series B in Q3 2022)



Sector: Cross-Industry Applications – HR^(a)

Geography: United States

Company Description: Phenom's Al-powered talent relationship management platform helps companies hire faster, develop

better and retain employees longer, ensuring that organizations are well-positioned to grow.

AVP Experience: US Growth team led Phenom's Series B in 2018 (follow-on investments in Series C and Series D)



IND¦C⊙

Sector: Verticalized Solutions – Finance & Insurance

Geography: Israel

Company Description: ThetaRay's Al-powered SaaS AML transaction monitoring and screening solution enables fintechs, banks and regulators to embed trust in cross-border and domestic payments while driving financial growth.

AVP Experience: EU Growth team evaluated ThetaRay during its Series C fundraise in Q1 2023.



Company Description: Indico uses AI and ML technology to automate the intake and understanding of unstructured documents,

emails, images, videos and audio files.

AVP Experience: US Early Stage team evaluated Indico during its Series B fundraise in Q4 2022.

Source: Pitchbook.

Note: Pitchbook screen follows same methodology as described in the Basis of Presentation. Not all ~1,000 companies are included on the market map on the prior page.

(a) HR is not included as a designated segment on the market map, however, there are likely several companies within the space that could be of interest to AVP.

The Al Investment Diligence Checklist

For Al companies in the application layer, there are multiple factors to consider when evaluating differentiation

Speed of Iteration & Updates

What is considered state-of-the-art in AI is changing at a rapid pace; companies need to stay up to date with the latest technology, architecture and models in order to offer a differentiated solution

Product & Customer Experience

With access to Al becoming increasingly accessible via OSS and private APIs, differentiation will increasingly be found in the product and customer experience itself

Enterprise Grade Standards

Data privacy standards and regulatory compliance are non-negotiable when it comes to Al-enabled applications looking to sell to enterprise clients

Commitment to Responsible AI (ESG)

When incorporating AI into SaaS applications, companies should ensure that steps are taken to reduce bias, increase transparency and uphold ethical standards

Technical Founders

Al is complicated – while many companies in the application layer won't be building their own Al from scratch, the technical sophistication and experience of founders will still be important to consider

Access to Proprietary Data

LLMs are trained on billions of parameters, but are not specific to any given company or use case – without proprietary, private training data, companies cannot refine these models to suit their exact product / use case

Private vs. Open Source Architecture

Private solutions (e.g., OpenAl's API) may be preferable over open source software (e.g., HuggingFace) as private solutions do not require continued maintenance, have higher levels of security and offer better support

Quantifiable ROI & Solution Relevance

It is important to understand how acutely a solution is solving a specific problem – market-leading solutions will explicitly reduce operating costs and increase worker efficiency / accuracy within a given niche

Multi-Platform

The ability to support or easily switch between multiple model platforms (e.g., Anthropic, OpenAl, Stable Diffusion) is important as different enterprise customers may require varying standards

Data Feedback Loop & Network Effects

Model performance and accuracy will rapidly increase if there is a flywheel effect (e.g., real-time user engagement data is collected and used to further train / finetune models, more data is captured as users increase)

6

8

10



Energy Consumption & Climate Change

Al requires extensive energy to function; however, there are positive applications of Al which may help to lower our carbon footprint

Benefits

- Despite its massive energy requirements, Al does have practical applications that can help to lower our carbon footprint
 - It's a bit of a catch-22, but AI itself can analyze mass amounts of complex climate change data to help us better understand what is happening in the world around us and inform policy decisions that address specific problems
- 87% of private and public sector CEOs believe Al is an essential tool in the fight against climate change
 - Moreover, 43% of organizations confirm having a vision for using Al in their own climate change efforts
- Based on research by BCG, AI can help to achieve a 5 10% reduction in global carbon emissions (2.6 – 5.3 gigatons)
 - Companies can use Al-powered data engineering to automatically track emissions throughout their carbon footprint
 - Predictive AI can forecast future emissions across a company's carbon footprint, in relation to current reduction efforts, new carbon reduction methodologies and future demand
 - By providing detailed insight into every aspect of the value chain, prescriptive AI and optimization can improve efficiency in production, transportation, and elsewhere, thereby reducing carbon emissions and cutting costs

Framework for Using AI to Combat Climate Change TOPICS Mitigation Adaptation and resilience **Fundamentals** Vulnerability and Reductions Hazard forecasting Measurement Removal exposure management Projecting regionalized Reducing GHG Climate research Environmental Macro-level emissions intensity long-term trends e.g., monitoring and modeling removal e.g., regionalized epidemics AND EXAMPLES measurement e.g., supply forecasting e.g., modeling of economic e.g., monitoring e.g., estimating modeling of sea-level and social transition for solar energy rise or extreme events remote carbon Strengthening forests and other such as wildfires and natural stock infrastructure Improving energy natural reserves floods Climate finance e.g., intelligent irrigation efficiency e.g., forecasting e.g., encouraging carbon prices Protecting populations behavioral change **Building early** Micro-level Technological e.g., predicting large-scale warning systems measurement migration patterns e.g., near-term Education, nudging e.g., calculating the Reducing e.g., assessing and behavioral chang prediction of extreme carbon footprint of greenhouse effects carbon-capture Preserving biodiversity e.g., recommendations events such as individual products e.g., accelerating aerosol e.g., identifying and cyclones for climate-friendly and chemistry research counting species Strengthen planning and Support collaborative Gather, complete, and Encourage climate-positive USES Optimize processes decision making Supply chain process data ecosystems behaviors Satellite and IoT data Policy and climate-risk analytics optimization Vertical data sharing Climate-weighted suggestions Filling gaps in temporally Modeling higher-order effects Simulation Enhanced Climate-friendly optimization and spatially sparse data Bionic management environments communication tools

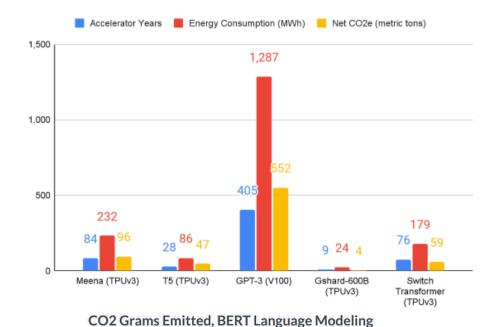
Source: Bloomberg, BCG and other publicly available information.

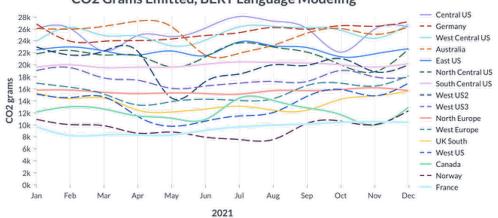
Energy Consumption & Climate Change (cont'd)

Al requires extensive energy to function; however, there are positive applications of Al which may help to lower our carbon footprint

Considerations

- As machine learning models have increased in scale, so too have the computational and energy requirements needed to create and sustain them
 - In 2020, computers consumed roughly 4-6% of the global electricity supply (up from 1-2% in 2018)
 - By 2030, this figure is projected to rise between 8-21%
- Taking OpenAl's GPT-3 model as an example, it took 1,287 MWh to train the model which generated a carbon footprint is 552 CO2e
 - For comparison, 1,287 MWh is comparable to powering 120 US homes in a year while 552 $\rm CO_2e$ is about as much as 110 US cars emit in a year
 - While training the model has significant upfront power costs, in some cases training is only 40% of the power burned by the model with the rest coming from billions of live requests to use the platform
- At the same time, the infrastructure required to power AI (i.e., data centers)
 has also accelerated
 - Between 2017 and 2020, energy consumption and carbon emissions associated with data centers doubled
 - Data centers are usually running near full utilization, meaning that a 20-megawatt facility would consume enough energy to power roughly 16,000 households
 - The location of a data center can greatly affect the amount of energy consumed; for example, a model trained in the Pacific Northwest would generate less carbon due to its use of clean



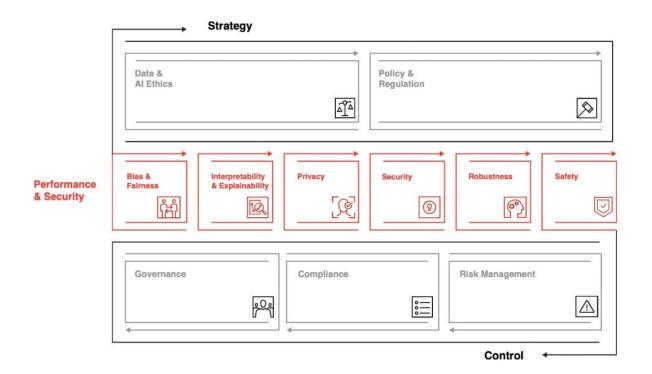


29

Responsible Al

Organizations scaling Al should be mindful of regulatory, privacy, security and ethical concerns

- Responsible Al is defined as the practice of designing, developing and deploying Al with the good intention to empower employees and businesses while fairly impacting customers and society
- Al is squarely on the minds of governments and regulators as an area of both promise and concern
 - In October 2022, the White House published its Blueprint for an Al Bill of Rights which included five main points
 - Automated systems should be safe and effective
 - Users should not experience algorithmic discrimination
 - Users should be protected from abusive data practices and have control over how their data is used
 - Users should know an automated system is being used and understand how it impacts them
 - Users should be able to opt out and talk to a human where possible
 - Similarly, the EU has proposed its own AI Act which would classify AI systems by risk and mandate various development and use requirements
- Organizations leading the push in Al, including Microsoft and Google, also have published (in 2017 / 2018) their own Al principles by which their systems will be measured



Source: Accenture, Google, Microsoft and PwC.



What's Next for AI?

Multi-Modal Al

To date, large language models have been just that – language. Companies like OpenAl are focused on multi-modal Al, or the idea that the input data can come in multiple forms, whether text, image or video. Through multi-modality, solutions including ChatGPT can become multi-purpose, combining skills in language and images to make the Al better and understanding both.

Algorithm Improvements & Next-Gen Hardware

With model improvements, experts contend that GPUs will pick up speed and remain an important part of the "computational power" variable in the formula that drives the next Al leaps. However, some Al hardware under development, such as neuromorphic chips or even quantum computing systems, could factor into the new equation for Al innovation.

Powerful Verticalized Applications

Within healthcare, for example, Al will allow for much more personalized medicine and bring a revolution in the use of large medical datasets. We'll see patient-specific treatments—for example, ones created using your genomic and expression data, which are much more likely to work.

Closing the Gap Between Al and Human Understanding

Deep learning continues to generate useful applications, but considerable work remains at the analytical level to understand how human cognition works in supporting problem-solving and critical thinking and creativity.

Efficient AI Models

Over the past five years, the following statement has largely held true – the larger the model, the more accurate the Al will be; however, a key breakthrough will occur when an extremely small dataset is proven to train Al systems as well as large datasets. Researchers are pushing to figure out ways to train systems on less data and are confident they'll find a viable solution. As a result, Al experts expect the "data" variable in the Al growth equation to flip on its head, with small datasets overtaking big data as drivers of new Al innovation.

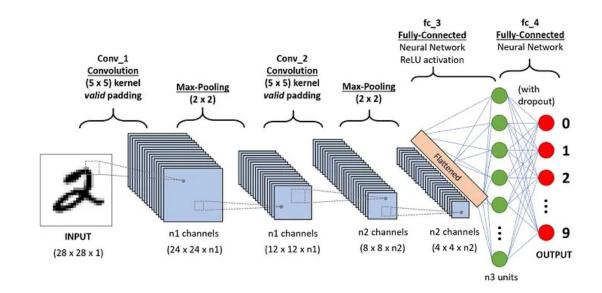
Unsupervised Learning

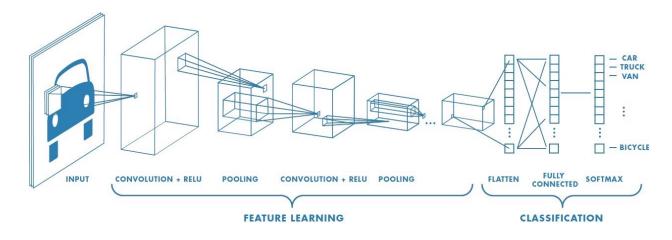
Unsupervised learning is the idea that AI systems are able to learn without human guidance or labels on the data that is fed into the systems to train them. Experts contend that we're still a long way away from complete unsupervised learning, but that the next wave of AI innovation will likely be fueled by deep learning models trained using a method somewhere between supervised and unsupervised learning.

Appendix: Al Architecture & Techniques

Convolutional Neural Networks (CNNs) - 1980s

- CNNs are deep learning algorithms that can take in an input image, assign importance (learnable weights and biases) to various aspects / objects in the image and be able to differentiate one from the other
 - From a use case perspective, CNNs are primarily used for image and video recognition tasks, including image classification, object detection, face recognition, medical imaging and natural language processing
- CNNs have three main types of layers, which are the Convolutional Layer, Pooling Layer and Fully-Connected (FC) Layers
 - Convolutional Layer: Core building block of a CNN, and it is where the majority of computation occurs
 - After each convolution operation, a CNN applies a Rectified Linear Unit (ReLU) transformation which allows for faster and more effective training by mapping negative values to zero and maintaining positive values
 - Pooling Layer: Conducts dimensionality reduction, reducing the number of parameters in the input
 - FC Layer: Performs the task of classification based on the features extracted through the previous layers and their different filters

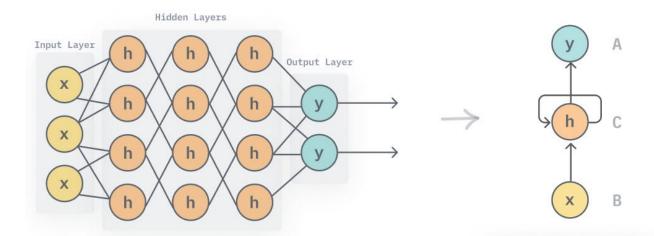


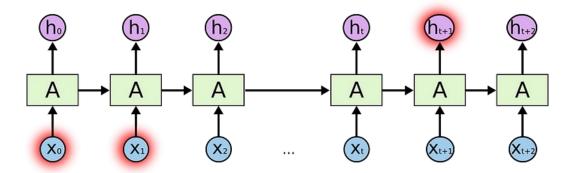


Source: IBM and other publicly available information.

Recurrent Neural Networks (RNNs) – 1986

- RNNs are a type of artificial neural network which use sequential data or time series data
 - These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (NLP), speech recognition, and image captioning
- Like convolutional neural networks (CNNs), recurrent neural networks utilize training data to learn
 - They are distinguished by their "memory" as they take information from prior inputs to influence the current input and output
 - While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of recurrent neural networks depend on the prior elements within the sequence
- Recurrent neural networks leverage backpropagation through time (BPTT) algorithm to determine the gradients, which is slightly different from traditional backpropagation as it is specific to sequence data
 - Through this process, RNNs tend to run into two problems, known as exploding gradients and vanishing gradients
 - One solution to these issues is to reduce the number of hidden layers within the neural network, eliminating some of the complexity in the RNN model

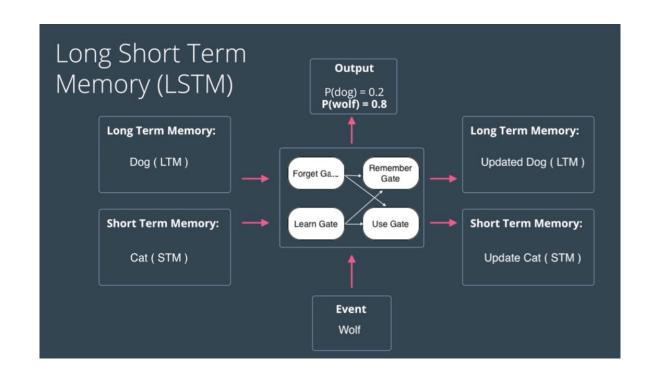




Source: IBM and other publicly available information.

LSTM (Long Short-Term Memory) – 1997

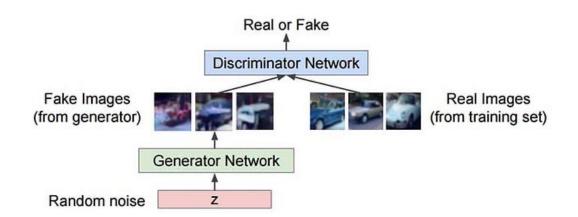
- LSTM is a popular type of RNN architecture which was introduced as a potential solution to the vanishing gradient problem
 - Vanishing Gradient Problem: When the gradient (i.e., a derivative of a function that has more than one input variable) is too small, it continues to become smaller, updating the weight parameters until they become insignificant (e.g., when this occurs, the algorithm is no longer learning)
 - The LSTM model works to solve the issue of long-term dependencies, or an RNN's inability to accurately predict the current state if the context to do so is well into the past
- For example, take the statement "Alice is allergic to nuts. She can't eat peanut butter."
 - The context of a nut allergy can help anticipate that peanut butter cannot be eaten because it contains nuts
 - However, if that context was a few sentences prior, then it would make it difficult, or even impossible, for the RNN to connect the information
- LSTMs have "cells" in the hidden layers of the neural network, which have three gates (input gate, output gate, forget gate)
 - These gates control the flow of information which is needed to predict the output in the network
 - For example, if gender pronouns, such as "she", was repeated multiple times in prior sentences, you may exclude that from the cell state

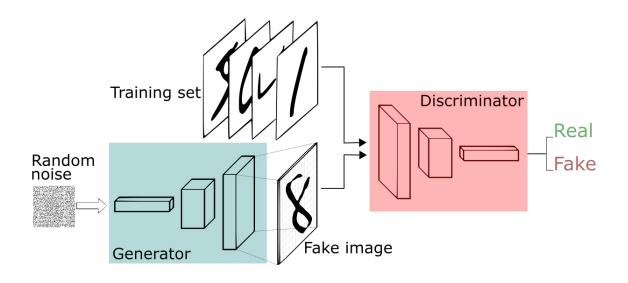


Source: IBM and other publicly available information.

Generative Adversarial Networks (GANs) – 2014

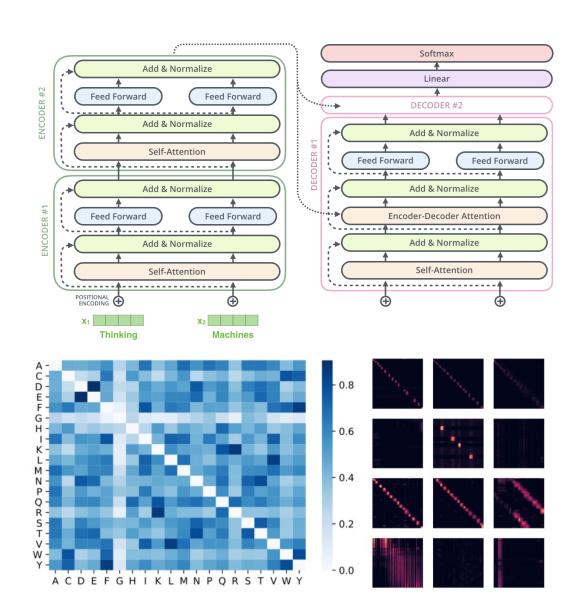
- GANs are algorithmic architectures that use two neural networks, pitting one against the other (thus the "adversarial") in order to generate new, synthetic instances of data that can pass for real data
 - First described in a paper in 2014, GANs are used widely in image generation, video generation and voice generation
- One neural network (generator) generates new data instances, while the other (discriminator) evaluates them for authenticity (i.e., the discriminator decides whether each instance of data that it reviews belongs to the actual training dataset or not)
 - Both networks are trying to optimize a different and opposing objective function, or loss function, in a zero-sum game – as the discriminator changes its behavior, so does the generator creating a double feedback loop
- However, there are several challenges when working with GANs
 - Given that GANs involve two competing neural networks, there is double the amount of complexity vs. just training one neural network
 - One network may overpower the other, such that neither can learn anymore
 - GANs can also suffer from "mode collapse", or when the generator only learns a small subset of the possible realistic models





Transformers - 2017

- Transformer modes are neural networks that learn context and thus meaning by tracking relationships in sequential data (like the words in this sentence)
 - First described in a 2017 paper from Google, transformers are among the newest and one of the most powerful classes of models invented to date
- Transformer models apply an evolving set of mathematical techniques, called attention or self-attention, to detect subtle ways even distant data elements in a series influence and depend on each other
 - Transformers use positional encoders to tag data elements coming in and out of the network
 - Attention units follow these tags, calculating a kind of algebraic map of how each element relates to the others
 - Attention queries are typically executed in parallel by calculating a matrix of equations in what's called multiheaded attention
- Transformers are in many cases replacing CNNs and RNNs for NLP
 - Before transformers arrived, users had to train neural networks with large, labeled datasets that were costly and time-consuming to produce
 - By finding patterns between elements mathematically, transformers eliminate that need, making available the trillions of images and petabytes of text data on the web and in corporate databases



Source: Nvidia and other publicly available information.

Transformers – 2017 (cont'd)

Input Embeddings

- Embeddings are numerical representations (i.e., vectors) of pieces of information
- The representation captures the semantic meaning of what is being embedded and helps ML models understand relationships and similarities between words (e.g., hotel and motel have similar meaning)

Positional Encoding

- Describes the location of a piece of data in a sequence, storing information about word order in the data itself (vs. the network)
- By encoding data, the neural network learns the importance of word order from the data

Multi-Head Attention

- The <u>attention mechanism</u> allows a neural network to look at every single word in an input sentence at the same time before generating an output sentence (vs. reading each word sequentially in one direction)
- <u>Multi-head attention</u> in the encoder specifically uses a type of attention called <u>self-attention</u> which relates different positions of a single sequence in order to compute a representation of the sequence (i.e., self-attention allows a neural network to understand a word in the context of the words around it)

Self-Attention Example

- Take the following to sentences: 1) "What book are you reading?" and 2) "Did you book the appointment?"
- When looking at each sentence, it is clear to us that the word "book" has a differing meaning given the broader context surrounding it
- Using the self-attention mechanism, Transformer models can recognize that "book" is being used in two separate parts of speech by attending to "reading" and "appointment"

<u>Decoder</u>

- The decoder takes the vectors (text converted to vectors by the encoder) and predicts the next word in the output sequence based on the context learned from the encoder and the previously generated words in the output sequence (this process occurs until it has generated the entire output sequence)
- The decoder is trained using a process called backpropagation, which adjusts the model's parameters to minimize the difference between the predicted output sequence and the actual output sequence

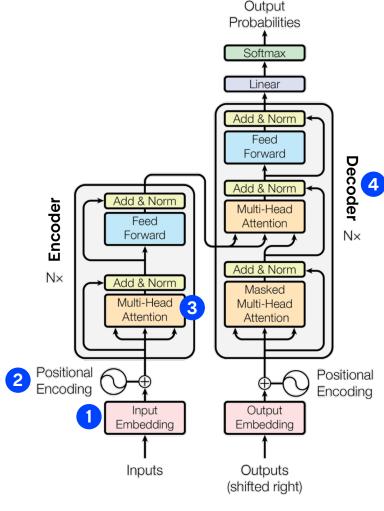
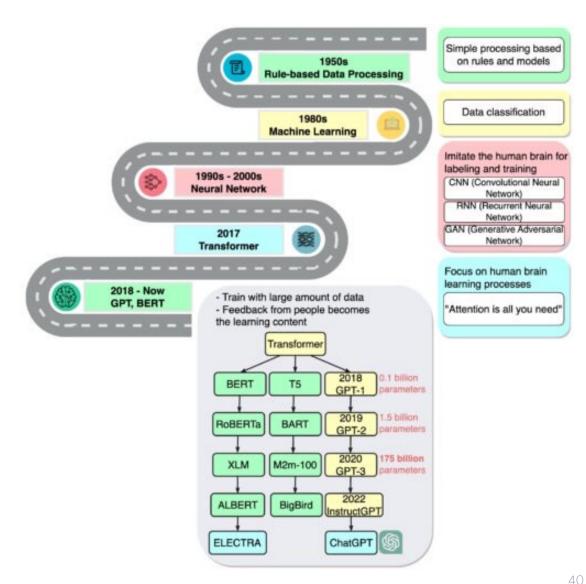


Figure 1: The Transformer - model architecture.

Foundation Models – 2018 to present

- A foundation model is an Al neural network trained on mountains of raw data, generally with unsupervised learning — that can be adapted to accomplish a broad range of tasks
 - Foundation models generally learn from unlabeled datasets, saving the time and expense of manually describing each item in massive collections.
- With the previous generation of AI techniques, if you wanted to build an AI model that could summarize bodies of text for you, you'd need tens of thousands of labeled examples just for the summarization use case
 - With a pre-trained foundation model, we can reduce labeled data requirements dramatically
 - First, we could fine-tune it to domain-specific unlabeled data to create a domain-specific foundation model
 - Then, using a much smaller amount of labeled data, potentially just a thousand labeled examples, we can train a model for summarization
 - The domain-specific foundation model can be used for many tasks as opposed to the previous technologies that required building models from scratch in each use case
- Since the advent of the Transformer, there has been a rush to produce larger and larger foundation models (LLMs), including OpenAl's GPT and Google's BERT



Source: Nvidia, IBM and other publicly available information.

Appendix: Overview of Generative Al

Overview of Generative Al

- Generative AI refers to a category of artificial intelligence algorithms that generate new outputs based on the data they have been trained on
 - Unlike traditional Al systems that are designed to recognize patterns and make predictions, generative Al creates new content in the form of images, text and audio
- Generative Al uses a type of deep learning called generative adversarial networks (GANs) to create new content
 - A GAN consists of two neural networks: a generator that creates new data and a discriminator that evaluates the data
 - The generator and discriminator work together, with the generator improving its outputs based on the feedback it receives from the discriminator until it generates content that is indistinguishable from real data
- Generative AI has a wide range of applications, including:
 - Images: Generative AI can create new images based on existing ones, such as creating a new portrait based on a person's face or a new landscape based on existing scenery
 - <u>Text:</u> Generative AI can be used to write news articles, poetry, and even scripts. It can also be used to translate text from one language to another
 - Audio: Generative AI can generate new music tracks, sound effects, and even voice acting

The Generative AI Tech Stack Applications User feedback mechanism Otome MLOps User databases Mosaic Foundation models co:here **Generative Architectures** Transformers Diffusion VAEs GANs

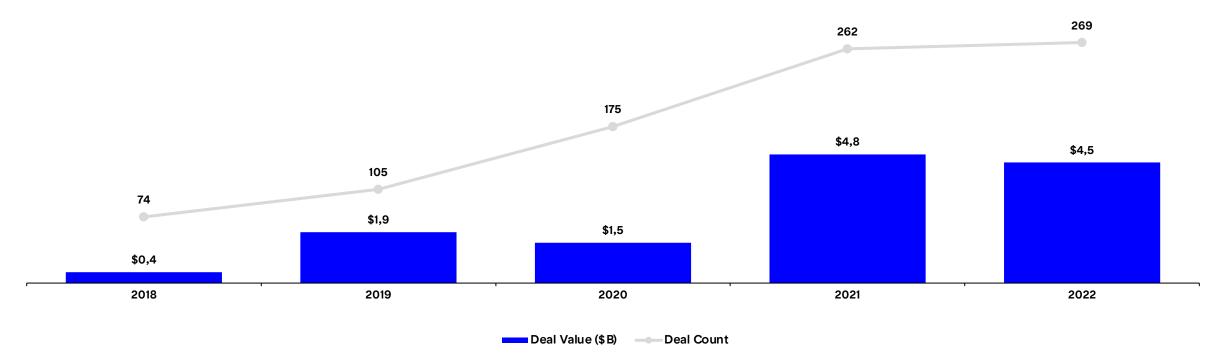
Source: World Economic Forum and Pitchbook.

Industry Drivers

- 1 The cost of foundation model training has been decreasing
 - While GPT-3 was costly to train (estimates placed a single training run at \$10 million), newer solutions including Stable Diffusion have brought the cost for state-of-the-art models down to ~\$600k
 - Continued advancements in hardware (e.g., Nvidia's H100 Tensor Core GPU) will continue to decrease the cost of compute
- 2 The cost to use foundation models by startups remains accessible
 - The cost of a typical response from an API (such as those provided by OpenAI, Cohere and Midjourney) are priced competitively to enable startups to
 launch applications on top of their software
- 3 CIOs are pushing for AI adoption across their business departments
 - 60% of CIOs plan for AI to gain widespread adoption across departments by 2025
- 4 Al-enabled productivity increases can help to solve labor shortages and cut costs
 - Al-powered tools, such as chatbots, assistants and copilots, can automate time-consuming tasks, freeing employees to focus on higher-level work and increasing productivity per worker
- 5 Al advancements have increasingly shifted from research to industry
 - Al researchers have steadily moved from academia to industry as opportunities have continued to grow in the private sector, leading to more companies focused on practical Al innovation
- 6 Increasing virality of consumer-facing products
 - As evidenced by ChatGPT, which reached 100 million MAUs faster than any technology within its first two months, consumer-facing AI products have the
 potential to go viral quickly upon launch

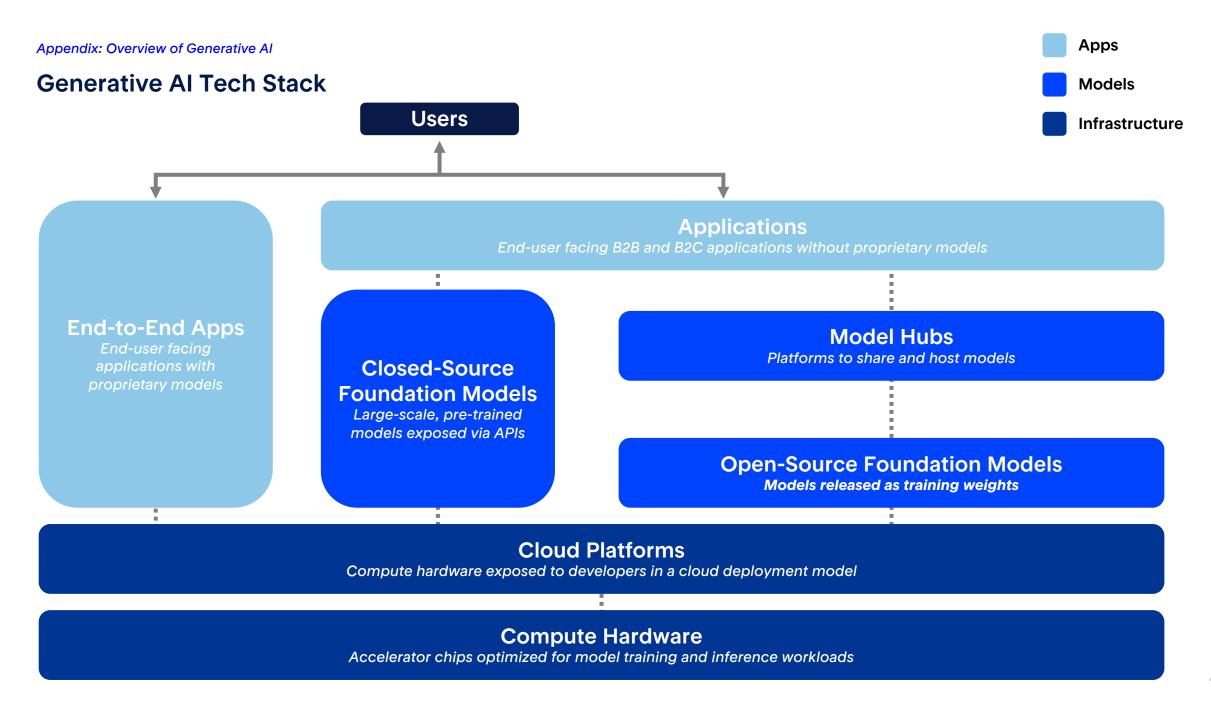
Source: Pitchbook.

VC Activity in Generative AI



	Top Generative AI VC Investors (since 2018)								
Investor	Deal Count	Seed	Early Stage VC	Late Stage VC	Venture Growth				
a16z	21	5	9	5	2				
Tiger Global	20	1	9	8	2				
Sequoia	20	5	12	3					
Amplify Partners	17	5	9	3					
Khosla Ventures	17	5	6	6					
Bloomberg Beta	15	4	8	3					
Index Ventures	15	4	9	ſ	1				
Soma Capital	15	11	3	1					
South Park Commons	14	5	6	3					
Alumni Ventures	14	7	5	2					

Source: Pitchbook and other publicly available information as of 2022.



Select Generative AI Applications & Use Cases

2D Media	Code	Vertical Applications
Avatars	Generation	Health
Content Suite	Documentation	Legal
Images	Web App Builders	Finance
Product Design	Audio	HR
Video	TTS / STT	Horizontal Applications
3D Media	Music	Cybersecurity
Object Synthesis	Editing	Sales & Marketing
Product Synthesis	Voice Dubbing	Customer Experience
Space Synthesis	Summarization	Search
Avatars & NPCs	Translation	Workflow Productivity

Key Generative AI Partnerships with Cloud Hyperscalers







aws stability.ai

Google & Anthropic

In February 2023, Google invested \$400 million in Anthropic, becoming its preferred cloud provider through its Google Cloud Platform

Partnership Overview:

- Anthropic will leverage Google's customdeveloped machine learning systems designed to run computationally intensive workloads
- Google Cloud intends to build large-scale, next-generation TPU and GPU clusters that Anthropic plans to use to train and deploy its AI systems

Microsoft & OpenAl

In January 2023, Microsoft invested an additional \$10 billion in OpenAl (OpenAl and Microsoft began their partnership in July 2019)

Partnership Overview:

- Microsoft will deploy OpenAl's models across its consumer and enterprise products, introducing new categories built on OpenAl's technology
- Azure will power all OpenAl workloads across research, products and API services
- Microsoft will increase its investments in development and deployment of supercomputing systems to accelerate OpenAl's research

AWS & Stability Al

In November 2022, Stability Al announced that it had selected AWS as its preferred cloud provider to build and scale its Al models

Partnership Overview:

- Stability Al uses Amazon's end-to-end machine learning service, SageMaker, as well as AWS' compute infrastructure and storage to accelerate its generative Al models
- Stability Al will also make its open-source models available on Amazon SageMaker JumpStart, the model hub of Amazon SageMaker, for all AWS customers

Exclusive access to best-in-class compute is critical for businesses building the infrastructure for LLMs

Source: Pitchbook and other publicly available information.

Appendix: Multiples Sensitivity by Year

Illustrative Al Investment Sensitivity – Entry Revenue Multiples (2019 investments only)

			Series	A		
(\$ in m	m)			ARR		
		\$1.0	\$1.5	\$2.0	\$2.5	\$3.0
>	\$17	17 x	12 x	9 x	7 x	6 x
ne	\$19	19 x	13 x	10 x	8 x	6 x
Pre-Money Valuation	\$22	22 x	14 x	11 x	9 x	7 x
re-	\$24	24 x	16 x	12 x	10 x	8 x
d	\$26	26 x	17 x	13 x	10 x	9 x

Series B								
(\$ in m	ım)			ARR				
		\$2.0	\$4.0	\$6.0	\$8.0	\$10.0		
>	\$52	26 x	13 x	9 x	6 x	5 x		
ion	\$58	29 x	15 x	10 x	7 x	6 x		
Pre-Money Valuation	\$65	32 x	16 x	11 x	8 x	6 x		
- Ja	\$71	36 x	18 x	12 x	9 x	7 x		
Φ _	\$78	39 x	19 x	13 x	10 x	8 x		

	Series C								
(\$ in m	m)			ARR					
		\$8.0	\$11.0	\$14.0	\$17.0	\$20.0			
>	\$289	36 x	26 x	21 x	17 x	14 x			
ne ion	\$325	41 x	30 x	23 x	19 x	16 x			
Pre-Money Valuation	\$361	45 x	33 x	26 x	21 x	18 x			
/a	\$397	50 x	36 x	28 x	23 x	20 x			
Δ -	\$433	54 x	39 x	31 x	25 x	22 x			

	Series D								
(\$ in mm) ARR									
		\$20.0	\$25.0	\$30.0	\$35.0	\$40.0			
> _	\$530	27 x	21 x	18 x	15 x	13 x			
ion	\$596	30 x	24 x	20 x	17 x	15 x			
Pre-Money Valuation	\$663	33 x	27 x	22 x	19 x	17 x			
re-	\$729	36 x	29 x	24 x	21 x	18 x			
Δ -	\$795	40 x	32 x	27 x	23 x	20 x			

Illustrative Al Investment Sensitivity – Entry Revenue Multiples (2020 investments only)

			Series	A		
(\$ in m	m)			ARR		
		\$1.0	\$1.5	\$2.0	\$2.5	\$3.0
>	\$23	23 x	15 x	11 x	9 x	8 x
ne	\$26	26 x	17 x	13 x	10 x	9 x
re-Money /aluation	\$29	29 x	19 x	14 x	11 x	10 x
re-	\$32	32 x	21 x	16 x	13 x	11 x
Δ -	\$34	34 x	23 x	17 x	14 x	11 x

			Series	В		
(\$ in m	m)			ARR		
		\$2.0	\$4.0	\$6.0	\$8.0	\$10.0
>	\$65	32 x	16 x	11 x	8 x	6 x
ion	\$73	36 x	18 x	12 x	9 x	7 x
Pre-Money Valuation	\$81	40 x	20 x	13 x	10 x	8 x
re-	\$89	44 x	22 x	15 x	11 x	9 x
Δ -	\$97	49 x	24 x	16 x	12 x	10 x

Series C								
(\$ in m	m)			ARR				
		\$8.0	\$11.0	\$14.0	\$17.0	\$20.0		
>	\$224	28 x	20 x	16 x	13 x	11 x		
ne	\$252	32 x	23 x	18 x	15 x	13 x		
Pre-Money Valuation	\$280	35 x	25 x	20 x	16 x	14 x		
re-	\$308	39 x	28 x	22 x	18 x	15 x		
Д _	\$336	42 x	31 x	24 x	20 x	17 x		

	Series D								
(\$ in m	nm)			ARR					
		\$20.0	\$25.0	\$30.0	\$35.0	\$40.0			
>	\$1,060	53 x	42 x	35 x	30 x	27 x			
ion	\$1,193	60 x	48 x	40 x	34 x	30 x			
Pre-Money Valuation	\$1,325	66 x	53 x	44 x	38 x	33 x			
	\$1,458	73 x	58 x	49 x	42 x	36 x			
Δ _	\$1,590	80 x	64 x	53 x	45 x	40 x			

Illustrative Al Investment Sensitivity – Entry Revenue Multiples (2021 investments only)

			Series	A		
(\$ in m	m)			ARR		
		\$1.0	\$1.5	\$2.0	\$2.5	\$3.0
>	\$29	29 x	19 x	14 x	12 x	10 x
ne	\$32	32 x	22 x	16 x	13 x	11 x
Pre-Money Valuation	\$36	36 x	24 x	18 x	14 x	12 x
re-	\$40	40 x	26 x	20 x	16 x	13 x
d	\$43	43 x	29 x	22 x	17 x	14 x

			Series	В		
(\$ in m	m)			ARR		
		\$2.0	\$4.0	\$6.0	\$8.0	\$10.0
>	\$116	58 x	29 x	19 x	15 x	12 x
ion	\$131	65 x	33 x	22 x	16 x	13 x
Pre-Money Valuation	\$145	73 x	36 x	24 x	18 x	15 x
- a a	\$160	80 x	40 x	27 x	20 x	16 x
Φ _	\$174	87 x	44 x	29 x	22 x	17 x

			Series	С		
(\$ in mm)				ARR		
		\$8.0	\$11.0	\$14.0	\$17.0	\$20.0
_	\$460	58 x	42 x	33 x	27 x	23 x
ne	\$518	65 x	47 x	37 x	30 x	26 x
Pre-Money Valuation	\$575	72 x	52 x	41 x	34 x	29 x
- Se	\$633	79 x	58 x	45 x	37 x	32 x
a	\$690	86 x	63 x	49 x	41 x	35 x

			Series	D		
(\$ in mm)				ARR		
		\$20.0	\$25.0	\$30.0	\$35.0	\$40.0
Pre-Money Valuation	\$800	40 x	32 x	27 x	23 x	20 x
	\$900	45 x	36 x	30 x	26 x	23 x
	\$1,000	50 x	40 x	33 x	29 x	25 x
	\$1,100	55 x	44 x	37 x	31 x	28 x
Δ _	\$1,200	60 x	48 x	40 x	34 x	30 x

Illustrative Al Investment Sensitivity – Entry Revenue Multiples (2022 investments only)

			Series	A		
(\$ in mm)				ARR		
		\$1.0	\$1.5	\$2.0	\$2.5	\$3.0
>	\$32	32 x	21 x	16 x	13 x	11 x
Pre-Money Valuation	\$36	36 x	24 x	18 x	14 x	12 x
	\$40	40 x	27 x	20 x	16 x	13 x
re-	\$44	44 x	29 x	22 x	18 x	15 x
d	\$48	48 x	32 x	24 x	19 x	16 x

			Series	В		
(\$ in mm)				ARR		
		\$2.0	\$4.0	\$6.0	\$8.0	\$10.0
>	\$118	59 x	30 x	20 x	15 x	12 x
ion	\$133	66 x	33 x	22 x	17 x	13 x
Pre-Money Valuation	\$148	74 x	37 x	25 x	18 x	15 x
-a -	\$162	81 x	41 x	27 x	20 x	16 x
T	\$177	89 x	44 x	30 x	22 x	18 x

			Series	С		
(\$ in mm)				ARR		
		\$8.0	\$11.0	\$14.0	\$17.0	\$20.0
_	\$290	36 x	26 x	21 x	17 x	15 x
ne ion	\$327	41 x	30 x	23 x	19 x	16 x
Pre-Money Valuation	\$363	45 x	33 x	26 x	21 x	18 x
Val	\$399	50 x	36 x	29 x	23 x	20 x
D	\$436	54 x	40 x	31 x	26 x	22 x

Series D								
(\$ in mm)				ARR				
		\$20.0	\$25.0	\$30.0	\$35.0	\$40.0		
Pre-Money Valuation	\$1,008	50 x	40 x	34 x	29 x	25 x		
	\$1,134	57 x	45 x	38 x	32 x	28 x		
	\$1,260	63 x	50 x	42 x	36 x	32 x		
	\$1,386	69 x	55 x	46 x	40 x	35 x		
Δ -	\$1,512	76 x	60 x	50 x	43 x	38 x		

AXA Venture Partners

www.axavp.com