

# The AI that Industries Can Use Today



# The AI That Industries Can Use Today

2023 will mark the year that AI moved to the forefront of public consciousness. In this cycle it is important to consider the how and why behind applying AI technology to solve critical problems.

## Introduction

AI algorithms have certainly taken a huge leap forward in the past decade, let alone in the past several months. While older technology is still often more effective for particular use cases, newer AI is no longer bound to follow the same scripted programming that produces determined outcomes. Instead, “Generative” AI programs can create seemingly new content. As with any cycle, less forward-thinking incumbents and hype-riding start-ups have now started scrambling, applying Generative AI and similar “new” technologies to their businesses without considering whether they have the right technology or application. Although Generative AI is hoarding the spotlight, the older AI approaches are often better solutions for core, American industries.

There are many acute pain points that AI has successfully addressed for the past decade, including increasing efficiency and alleviating bottlenecks. However, with AI exuberance swirling around nearly every industry, first principals are more important than ever and truly impactful companies outside of big tech will have capital efficient answers for data ownership, data comprehension, defensibility, and market pull. The most successful tech players addressing the real economy have already been implementing strategies to create a data moat, utilize AI to mine for insights, and invoke a business model that monetizes those insights.

## Background: Types of AI

The root of all AI is statistical modeling: fitting data to a curve and predicting where along that curve future data points will lie. Interpreting written text and creating a digital image is similar to the algebraic exercise of finding a line that best matches a set of points. AI uses trillions of points of data and can find and mimic intrinsic patterns within the data, but the underlying concept is similar. Over the past 70 years, AI has grown from simple algorithms like the ELIZA chatbot/therapist, to more generalizable AI like IBM’s Watson which

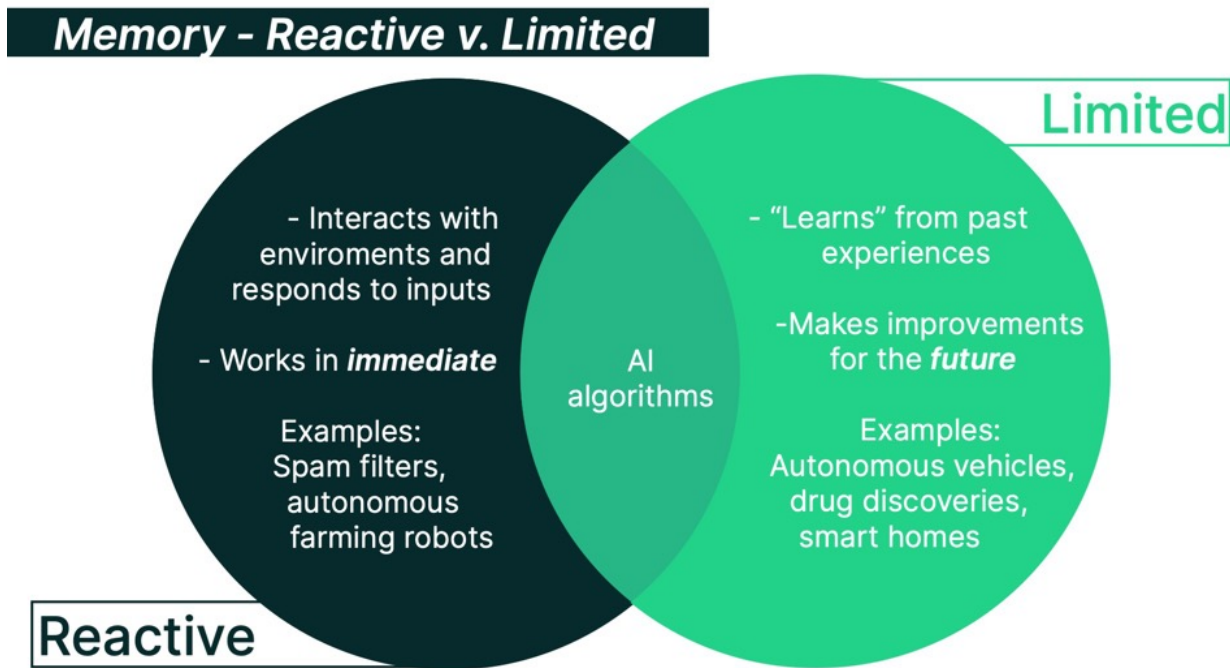


pivoted from winning Jeopardy to assisting in medical diagnoses and financial decisions.

Current AI algorithms fall broadly into two levels of complexity, both with tangible application: Reactive and Limited Memory.

### Reactive AI

Reactive AI algorithms interact with their environment and respond to current inputs. Most of what we consider AI today falls into this category: from [spam filters](#) to [autonomous farming robots](#). They can take in immediate data – a partially completed sentence, the state of a chess board, or soil conditions – and decide the best action to take. However, they have no understanding of the past or future or the context they are working in. For example, these algorithms can make decisions based on a set of parameters like “the soil nutrients are low, add fertilizer” but they cannot reason out “we will get a better yield next year if we leave this field fallow this year because we planted high-input crops here last year”. Deep Blue, the chess bot that defeated grandmaster Garry Kasparov in 1997, was a type of reactive AI.



Although this sounds limited, reactive AI is often better than more complex systems in terms of both efficacy and cost-savings. Reactive AI systems still benefit greatly from data and training to perform important tasks with little input from humans. able to capitalize on the new availability of high-speed internet, better content, and algorithms that connected people to what they wanted to watch.

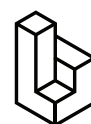
#### Limited Memory AI

In contrast, Limited Memory algorithms can benefit from past experiences to adjust their future actions and understand the context of new inputs. This gives them the ability not only to “learn” a fixed decision system but also to improve future decisions. Anything from medical image analyzers to language translators can fall under this category.

The advantage of Limited Memory AI is that it can incorporate historical data into its decision making and update parameters with new data without needing specific programming. A lesion detection algorithm can figure out which images contain an abnormality by guessing, checking whether it was correct, and updating its predictions for next time. This is immensely useful for autonomous vehicles, drug discovery, and smart homes.

#### Reactive Machines (“Dumb” AI)

Reactive machines have no memory-based functionality, exclusively executing automated tasks based on a limited input. Reactive machines are important to note due to their relationship with other AI. Since the early days of the technology, one constant is overrepresentation of differentiation, and a lack of distinction between the catch-all term of “AI” and the underlying technology. Oftentimes startups use terminology like “cutting edge AI” to describe a simple reactive machine (i.e., a business rule engine) without tech defensibility. Since most non-data-scientists don't have the sophistication to properly diligence the technology, this contributes to and compounds a hype cycle. However, these basic “AI” forms can have real value and many ML applications are similarly less defensible or differentiated than one might assume. This is why having a core foundation of strong data often outweighs a technology barrier and sophisticated engineering talent when evaluating this type of company.



## What is Modern AI?

### The Foundation: Machine Learning

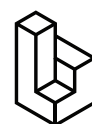
Machine learning (ML) is the core advancement that enables modern AI. This technique allows algorithms to “learn” from large datasets to accomplish a goal without being explicitly programmed for a specific task. ML is an iterative process where the algorithm is fed a large amount of reference data and makes guesses about classifying that data. It can then check its answers and adjust parameters and try again. After many iterations, ML algorithms can reach high levels of performance when making predictions, classifying objects, and making decisions.

ML is not new technology. One of the first use cases of ML was in detecting abnormalities in medical images. A single imaging test can produce thousands of images, and subtle abnormalities can be overlooked. Researchers use statistical techniques in ML to automatically identify patterns within these images and recommend follow-ups with the physician. The ingeniousness of ML is that it does not need to be programmed explicitly to identify cancers. Instead, by training on many radiological images, machine learning algorithms “learn” to recognize patterns and make data-driven decisions.

### Neural Networks, Natural Language Processing, and Beyond

Neural networks (NN) are an extension of consisting of layers of interconnected “neurons” which each recognize a piece of the overall structure and pass that analysis up to the next layer which recognizes a larger piece and so on. Through training, neural networks can automatically learn and adjust the weights and biases of their interconnected neurons, improving their ability to make predictions or decisions. This ability to adapt and generalize from data allows neural networks to tackle diverse and challenging tasks such as image and speech recognition, natural language processing, and even control systems.

Natural Language Processing (NLP) models can take this process a step further to enable computers to understand and interact with human language by analyzing and interpreting text or speech. These models can classify text, determine the sentiment of a statement, translate between languages, and more. Any application that requires human-like communication such as customer service, data analysis, and prompts for content generation, likely relies on NLP models.



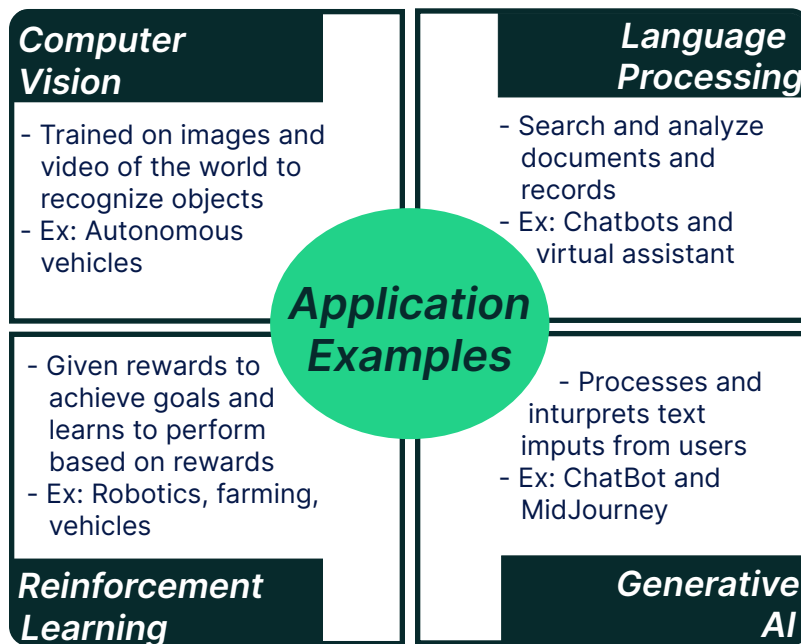


There are several other types of NNs, for example, recurrent neural networks (RNNs) which are commonly used in NLP tasks that involve sequential data, such as language modeling or translation because the models can capture relationships between words throughout a sentence. Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) are variations of RNNs that are better at identifying and retaining important information in the context of NLP. In addition to RNNs, other types of neural networks, such as convolutional neural networks (CNNs) and transformer models, are also employed in NLP tasks.

Each of these algorithms has its own specific specialties and limitations and it is critical to know when to use which one to optimize use of time and investment dollars.

### Applying AI to the Real World

The ability to generalize from examples and adapt to new information means that AI is extremely valuable across a number of diverse fields, from [risk prediction](#) systems for farmers to AI enhanced [robotics](#).



### Computer Vision Examples

Computer vision is one of the most often touted applications of AI. AI trained on images and video of the world can learn to recognize objects like vehicles or plants and understand how to perform its task in that environment.

There are many applications of computer vision in agriculture and manufacturing. [Vicarious](#), was one of the earliest players using computer vision to provide industrial robotic arms in a robots-as-a-service model. Using satellite imagery, companies such as [Reforestum](#) and [KoBold Metals](#) are similarly measuring [deforestation](#) and [searching](#) for the rare earth metals that will be needed to transition to a green economy. Another set of applications involve a wide range of anomaly detection from [QA](#) in manufacturing and [infrastructure](#) to [cancer](#).

### Natural Language Processing Examples

NLP is often effective playing a supportive role for transformational software platforms. For example, NLP enabled chatbots and virtual assistants can understand users and provide automated customer care. In addition, NLPs can search and analyze reference documents and medical records to generate useful analysis and recommendations.

[OpenEnvoy](#) uses NLP to digitize invoices and management and searches for inconsistencies and inefficiencies in logistics and other businesses.

### Reinforcement Learning Examples

Reinforcement Learning (RL) is a variation of ML where the algorithm is given a set of rewards for achieving a goal and learns to efficiently perform its task by iterating and adjusting its actions to maximize those rewards. This is another core technology for autonomous vehicles and is already in use on some [farms](#). The technology has similarly proven efficacy in more trivial, but similarly complex applications; in 2016 Google's DeepMind team used this type of algorithm to create [AlphaGo](#) to beat world champion go players.

Beyond winning games, is used extensively in the field of robotics with examples include [Robust.ai](#) using RL to map out efficient paths through warehouses for its robots to save workers time and effort and [Bright Robotics](#) applying similar techniques to automate electronics manufacturing and QA.



## Generative AI Examples

Generative AI has become wildly popular in the past year, with ChatGPT gaining over [100M](#) users within two months of launch. Built on a core foundation of traditional ML, generative systems like ChatGPT (the GPT stands for “generative pre-trained transformer”) and MidJourney build on earlier work in AI to go beyond making predictions; they can “generate” new content based on existing data and creative inputs.

These Generative AI models process and interpret text inputs from users using various AI components. The core advancement that enables these Generative AI models to efficiently learn patterns, relationships, and semantic structures in human language is the Transformer (detailed in Appendix A). The output is coherent and, importantly, contextually relevant responses and [content](#).

Some industries will completely change. These fears have already come to volition with [BuzzFeed replacing much of their staff with AI](#), and the Writers Guild of America [on strike](#) to avoid the same. Basic writing tasks, repetitive tasks, and applications with low sophistication broadly seem to bear risk of disruption from Generative AI.

Applications in core industries like farming and manufacturing are also emerging. In the industrials field, [Aether Biosciences](#) uses using Generative AI to formulate novel chemical compounds, saving many years of laboratory work. For healthcare patients, [Carbon Health](#) recently released [their AI-based electronic health records system](#) to write and store complete notes from patient appointments. Agriculture is benefitting as well; companies like [BankBarn](#) have dived into Generative AI; BankBarn uses generative models to forecast milk yield providing better underwriting and automation farm loan process, making sure farm owners have access to the cash they need.

## **The Generative AI Cart is Before the ML Horse**

While undoubtedly powerful and exciting to consumers, Generative AI has its limitations and questions:

### 1) *How are you going to get data?*

It is extraordinarily difficult to build Generative AI networks and to have them function at high levels. GPT-3 was trained on about 45 TB of data (roughly a quarter of the Library of Congress) at the cost of several million dollars. Obtaining this much data required the researchers to scrape a large portion of the internet, including Wikipedia, Reddit, and Twitter.



**builders | vc**



Image generators like StableDiffusion similarly collected data from public art sites like DeviantArt. There are several issues with this approach, the most basic being that the dataset is averaged. The algorithm will always be limited by the quality of data. There are further [questionable legality](#) of taking other people's artwork, writing, and creative content.

Eschewing public data, companies might design their own proprietary data sets focused on, for example, the agriculture or manufacturing industry. But this effort requires significant resources and potentially a consortium of partners prohibitive to all but the largest tech companies. Google's [Mineral](#) has compiled a formidable ecosystem including growers like Driscoll and industry leaders like Syngenta to gather to over six million images per plant species per day. This included the development of a proprietary [rover](#) for obtaining high quality images over hundreds-of-thousands of different environments. Mineral's generative elements feed into use-cases like determining the type of fungus or disease impacting a strawberry crop, given an individual fruit and pattern of fungal spread will never be exactly the same in nature. As these tech behemoths develop use-cases and their corresponding models they will [block others](#) from benefiting from the same information.

### 2) *How will the unit economics and capex costs pencil out?*

Processing this data is equally taxing and potentially prohibitive for start-up companies. Rumored costs for startups running proprietary LLMs have ranged from \$3-5 MM / month for leading companies in the Generative AI space, and it's estimated that OpenAI pays around [\\$700,000 a day to run ChatGPT](#), notwithstanding their high training costs. Similar to limitation (1), resource limitations for generating data and training models may self-select out all but established, large tech companies like Google and Microsoft.

### 3) *How do you hold on to data?*

Companies based on Generative AI have [limited defensibility](#). Relying on private databases isn't an option and it will be very difficult to differentiate companies using publicly available ones.

Data processing and algorithm training are major [rate-limiting](#) factors and even when they do become available, these databases will still face issues of data quality and consistency. Another approach is to specialize in designing ideal prompts for these industries, which could be [easily copied](#) or disrupted [others](#).



And critically, the United States does not allow copyrights on AI generated work so the outputs of these models might not have any protections once they are out in the world.

#### 4) *If you use these systems, can you rely on the data?*

Because ChatGPT is founded on an amalgamation of the Internet, it is constrained to the same biases and misunderstandings of the average internet user. While ChatGPT and others are great at confidently giving plausible answers, they are less good at giving correct answers. Already several prominent incidents have tarnished Generative AI's reputation including a [career-ending legal brief](#) and potential [defamation cases](#).

More broadly, Generative AI is not able to identify the source of its responses or determine whether they are factual. These results include the biases of their source material and can propagate [bigotry](#) or provide false information. As these models get better, it will be even harder to detect these issues.

### Final Thoughts

Since the early 2000s, companies have successfully leveraged types of AI including computer vision, NLP, and reinforcement learning to demonstrate meaningful impact across many fields. The demystification of AI due to the hype of Generative AI platforms like ChatGPT has already catalyzed interest in and potentially will speed up adoption for ROI-generating technologies in industries traditionally intimidated by everything Silicon Valley stands for. Those that stand to benefit the most are core American industries such as agriculture, manufacturing, logistics, real-assets, and healthcare. Even the most tech-resistant of customers in these sectors now accept that the incumbent way of doing things needs to be replaced by data driven decision-making and new distribution methods.

There is also reason to be cautiously optimistic for some Generative AI use-cases as companies aggregate large defensible datasets or determine unique and user-friendly applications. However, solving the inevitable resource constraints and big-tech competitive behavior will be significant challenges to those models. It seems many are throwing Generative AI at problems without thinking through how or why it will solve them.

At the end of the day, the heart of any AI, generative or not, is data. The optimal technique for gathering proprietary data is to serve as an industry's leading AI platform. Companies sticking to this mindset will win, with or without the flashy Generative AI exuberance of 2023, and cheers to that.



**builders | vc**

San Francisco  
601 California St, Suite 700  
San Francisco, CA 94108  
investments@builders.vc

[www.builders.vc](http://www.builders.vc)



**builders | vc**