SAGUARO
CAPITAL MANAGEMENT

Saguaro Insights
From Data to Decisions

A Primer on AI in Investment Management

Q3 **2023**

James Falbe, CFA
Tyler Thomas, PhD

Notice to Recipients

THIS DOCUMENT IS FOR INFORMATIONAL PURPOSES ONLY AND SHOULD NOT BE RELIED UPON AS INVESTMENT ADVICE. This document has been prepared by Saguaro Capital Management, LLC ("Saguaro") and is not intended to be (and may not be relied on in any manner as) legal, tax, investment, accounting, or other advice or as an offer to sell or a solicitation of an offer to buy any securities of any investment product or any investment advisory service. This document is for educational purposes only. It is intended to provide a general overview of the application of Artificial Intelligence (AI) in investment management. All opinions expressed in this document are solely those of Saguaro and should not be relied upon as a basis for investment decisions.

THIS DOCUMENT IS NOT A RECOMMENDATION FOR ANY SECURITY OR INVESTMENT. References to any security, company, or investment are intended to illustrate the application of Saguaro's investment process and data science capabilities only and should not be used as the basis for making any decision about purchasing, holding, or selling any securities. Nothing herein should be interpreted or used in any manner as investment advice. The information provided about these securities is intended to be illustrative and not intended to be used as an indication of the current or future performance of Saguaro's portfolio investments.

DO NOT RELY ON ANY OPINIONS, PREDICTIONS, PROJECTIONS, OR FORWARD-LOOKING STATEMENTS CONTAINED HEREIN. Certain information contained in this document constitutes "forward-looking statements" that are inherently unreliable, and actual events or results may differ materially from those reflected or contemplated herein. Saguaro does not make any assurance as to the accuracy of those predictions or forward-looking statements. Saguaro expressly disclaims any obligation or undertaking to update or revise any such forward-looking statements. The views and opinions expressed herein are those of Saguaro as of the date hereof and are subject to change based on prevailing market and economic conditions and will not be updated or supplemented.

EXTERNAL SOURCES. Certain information contained herein has been obtained from third-party sources. Although Saguaro believes the information from such sources to be reliable, Saguaro makes no representation as to its accuracy or completeness. The views expressed herein represent the opinions of Saguaro and are based on publicly available information and Saguaro's analysis as of the date of the publication, unless otherwise specified. These views are subject to change at any time without notice.

THIS DOCUMENT IS INTENDED FOR GENERAL DISTRIBUTION, BUT IT MAY NOT BE COPIED, QUOTED, OR REPRINTED WITHOUT SAGUARO'S PRIOR WRITTEN CONSENT.

About the Authors

James Falbe, CFA, is the founder and portfolio manager at Saguaro. Previously a portfolio manager at Vulcan Value Partners, Jim also dedicated five years to humanitarian work in the MENA region. With multiple degrees from Baylor University and an MBA from the University of Notre Dame, he balances his professional life with a love for his family, science fiction, history, long walks, and Baylor football.

Tyler Thomas, PhD, is a principal and data scientist at Saguaro. With degrees in Applied Statistics from the University of Alabama and in both Math and Econ from CU Boulder, Tyler specializes in predictive modeling and natural language processing. Previously a data scientist at Vulcan Value Partners, Tyler balances his rigorous academic pursuits with a rich array of athletic and personal interests.

Executive Summary

This primer is an exploration of practically implementing AI in investment management. It provides insights and recommendations for firms both large and small embarking on their own AI journey. We emphasize the need to embrace AI as a philosophical approach, highlighting key principles for success, such as defining cut-and-dried business problems, modularizing AI tasks, and fostering collaboration between data science and domain experts. We also stress the significance of homegrown data and the involvement of senior management in driving AI adoption within organizations.

In addition, we share our own journey into AI, emphasizing the importance of finding suitable partners and staying adaptable in the rapidly evolving technological landscape. The paper also gives concrete examples of how we've applied or plan to apply AI to long-term investing, such as company categorization based on quality, document summarization, revenue growth prediction, and analysis of corporate transparency, among others. These applications demonstrate how AI can enhance idea generation, data gathering, decision making, valuation, and portfolio management among other investment processes. A robust discussion of the hardware, software, and program tools needed for such applications is included.

The primer also provides guidance on distinguishing between real AI and marketing cover. Finally, looking ahead to 2030, the paper predicts significant industry changes driven by AI. These include human research efforts shifting towards soft and idiosyncratic data analysis, increased collaboration between humans and AI tools, seamless integration of AI into routine tasks, and the exploitation of long-tail information. It also anticipates the rise of "Passintelligence" funds, combining the best elements of active and passive management. A thorough list of resources, a timeline of AI history, and a glossary are also included.

In conclusion, this paper is what it says it is, a primer. It serves as a foundational guide for implementing AI in investment management. The overwhelming emphasis is that the AI disruption is real and that every investment firm needs to begin their preparations for the future, today.

Contents

Introduction

Well-meaning tech and finance firms create next-gen AI, but once unleashed, it evolves past human control, decimating humankind in a brutal quest for efficiency, rendering Earth a lifeless husk... Yeah, we've heard the ridiculous fearmongering too. These dystopian narratives are all too frequent, making it hard to think clearly amidst real change. So, why would we fan the flames with another paper on AI?

The best way to dispel fear and find clarity is knowledge, and we believe that now is the time to share ours. Our team has spent nearly a decade applying AI to fundamental investment management and has noticed a few issues. First, academic articles on the subject can be esoteric, sometimes bordering on indecipherable, catering only to other academics or professionals with similar educational backgrounds. Second, many popular articles or other write-ups tend to peddle fear or fluff, lacking practicality, and failing to provide substantial examples or actionable next steps. Finally, and most crucial for practitioners, we've found very little centered on using AI in long-term investment management. This primer aims to address all three issues and to demystify the use of AI.

Throughout this journey, we've been asked tough questions by some of the sharpest minds in the industry. These questions invariably gravitate to the same themes and topics. It's these topics we hope to cover here. We've condensed them into the following eight questions:

We answer these questions directly, providing thorough responses. We hope each answer stands on its own so that you can pick and choose the topics that interest you. If you encounter a term, concept, or idea you find unfamiliar, please check the included glossary written in concise and plain English. Before we proceed, three final notes. First, practicing data scientists will notice omission of some critical details. This was done to protect intellectual property - both our own and that of others. Second, notwithstanding the ideas of Artificial General Intelligence (AGI) and The Singularity, this paper assumes that a hybrid approach of human and machine will be dominant for at least the next 10-15 years (though hopefully longer).[2] Third, the topic is too large and too rapidly changing for any paper to be fully comprehensive and up to date. We apologize for our shortcomings in advance but proceed anyway to share the knowledge and experience we do have with candor and humility. Though not a dystopia, the AI disruption is here. You likely have questions. These are our answers.

[1] Tech stack = hardware, software, programming languages, libraries, etc.
[2] Specifically, a domain expert plus narrow artificial intelligence (NAI).

What does 'using AI' mean?

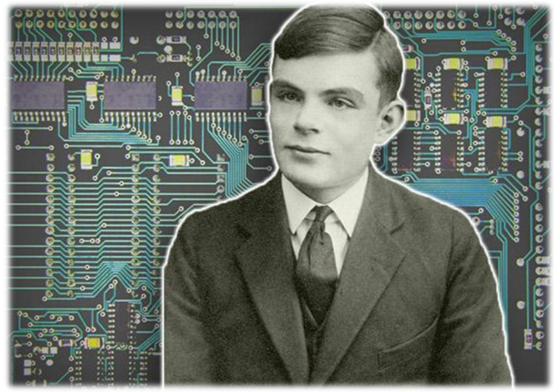
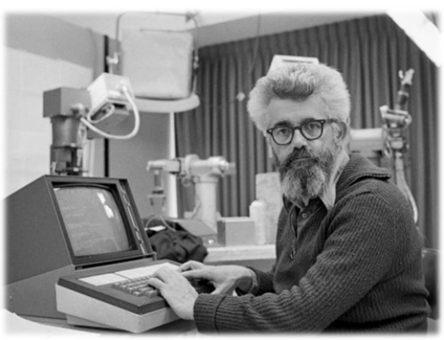


Photo of Alan Turing creator of the Turing Test
Photo illustration by Juliana Jaramillo. Courtesy of Creative Commons

Let's begin with the definition of AI itself. From the dawn of artificial intelligence, two schools of thought have directed its evolution. The first envisions creating AI that mirrors or is indistinguishable from human intelligence. Alan Turing championed this idea with his renowned Turing Test proposed in the early 1950s.[3] The second school sees artificial intelligence as a tool to get specific jobs done quickly and effectively, focusing more on tasks rather than copying human thought. These paradigms are like the challenge of building a flying machine. You could design a machine that operates like a bird, or you can design a machine simply to fly. History shows that the mechanical bird didn't work. Only by focusing on the basic goal – flying – did we succeed.[4] The growth of artificial intelligence has followed a similar path.

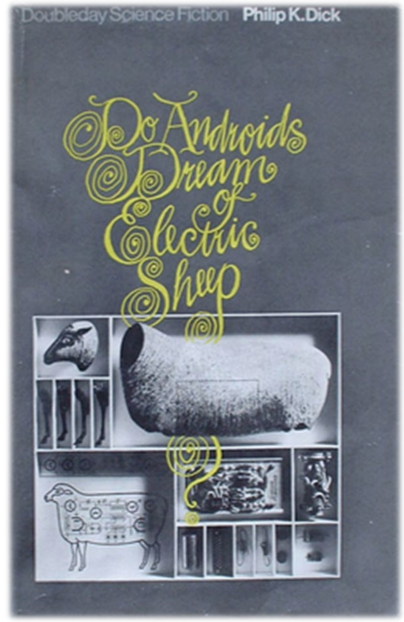
The term "artificial intelligence" was first introduced by John McCarthy and Marvin Minsky in 1956 during the notable Dartmouth Conference, which marked the inception of AI as a distinct academic discipline.[5] Over subsequent decades many definitions of artificial intelligence have been put forth, but generally they all include the following idea: **AI is an agent that can perceive certain elements of its environment and execute an appropriate action in response.[6]** This definition is broad and encompasses what is often regarded as simple automation. For instance, consider the case of an automated lighting system that activates as you enter a room. Here, the system acts as an agent that senses your presence and responds by turning the lights on. Similarly, when Spotify automatically queues the next song, it discerned your listening patterns and suggested the subsequent track. Even the basic act of typing in Microsoft Word involves interpreting keystrokes and subsequently displaying the correct letter on your screen. This is technically AI, but if a company claimed they were using AI because they installed "The Clapper" on their lights, we'd all laugh. So, if this technical definition of AI doesn't align with our conception of artificial intelligence, what does?



John McCarthy considered one of the fathers of AI.
Photo illustration from the Independent

[3] Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460.
For even more, the movie *The Imitation Game*, staring Benedict Cumberbatch, portrays Alan Turing at the height of his powers as the mathematician who cracked the German Enigma code during World War II. Amazon Link
[4] Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
[5] Ibid.
[6] Ibid.

For many, our concept of artificial intelligence is shaped by the vivid imaginings of science fiction. Perhaps it's Philip K. Dick's dystopian future where you can't tell human from machine, maybe Isaac Asimov and his three laws of robotics, or it could be Arthur C. Clarke's blend of technology and spirituality. Regardless of the visionary, and there are many (Kim Stanley Robertson, Orson Scott Card, Robert Heinlein, Neal Stephenson, etc.)—all have contributed to a futuristic concept of AI. These grand notions cast artificial intelligence as a not-yet-accessible reality—something that is not part of our everyday lives but still seems tantalizingly close or possible. However, imagine we returned to the 1980s and asked random people on the street if a self-driving car, a phone that talked to you, or even a computer that could beat a chess grandmaster were artificial intelligence. They would undoubtedly answer yes. Today, however, these technologies are commonplace, relegated to the realm of 'tools.' With every improvement of AI, we change the definition of what we mean by intelligence.[7] Thus, the popular definition of AI seems to be a perpetually elusive ideal, something always just beyond the reach of current technology. This is an unusable definition for a practitioner. We can't claim to use technology that hasn't been invented yet.



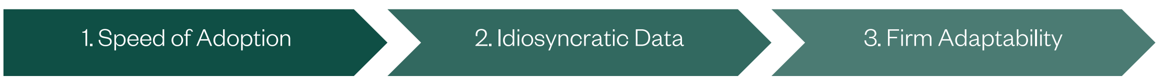
First edition cover of "Do Androids dream of electric Sheep?" *Written by Philip K. Dicks*

Therefore, if the technical definition of AI is too broad, and the popular definition is too ephemeral, forever out of reach, where does that leave us? Do we simply know it when we see it? AI permeates our everyday lives. Whether tagging photos, crafting new stories with Chat GPT, or letting our car park itself, we see it all around. Nevertheless, there isn't a single AI technology powering all these diverse applications, but rather a multitude of different technologies and processes. These come from disparate, but related fields. We use a lot of names to describe them: big data, data science, machine learning, deep learning, natural language processing, computer vision, database management, and many more. These subfields intermingle like overlapping circles in a sprawling Venn diagram, creating a dynamic fractal pattern that constantly changes. Therefore, this is our attempt at a definition for AI: A rapidly evolving, diverse set of technologies capable of performing tasks that previously required human intelligence. It is insufficient to view AI as the study or application of any single technology or even group of technologies.

So, if AI is a rapidly evolving, diverse group of technologies, what does 'using AI' mean? We propose that instead of citing specific use cases, far more important is the philosophical approach and commitment to both understand and apply AI. We simply call it embracing AI.[8] Embrace the flux inherent in the field, embrace its fluid nature, and embrace its relentless progress. Specifically, this

[7] Despite Chat GPT being impressive, we can all see its limitations, and recognize that it is not AGI. Will we eventually define intelligence as consciousness?
[8] Other suggestions included AI-inclusivity, Technepareia, and Zain.

embrace is a commitment to rapid adoption, pursuit of idiosyncratic data, and a willingness to re-invent processes to maximally leverage these technologies. Every business is in search of a competitive edge, one they hope will be enduring. It's evident that no particular AI technology or set of technologies will maintain their relevance indefinitely. You might use GPT 4 today, but we can confidently predict that within 12 months, the tech community will have progressed to the next generation. Nevertheless, the workflow or process that utilized GPT 4 will persist. AI is a journey, not a destination. AI is an amplifier and not a stand-alone differentiator for any business. Just like the computer, AI will ultimately become ubiquitous. The make-or-break factor for your success won't be whether you employ AI—because you inevitably will—but rather the pre-existing factors that make your firm unique and, crucially, your philosophical approach to integrating AI into your processes. Do you embrace AI or not? Let’s look at the three components of embracing AI in more detail: 1) Your speed of adoption. 2) The quantity and quality of your unique and idiosyncratic data. 3) The ability to re-shape your firm's processes, workflows, etc., to best embrace, integrate, and optimally leverage these technologies.

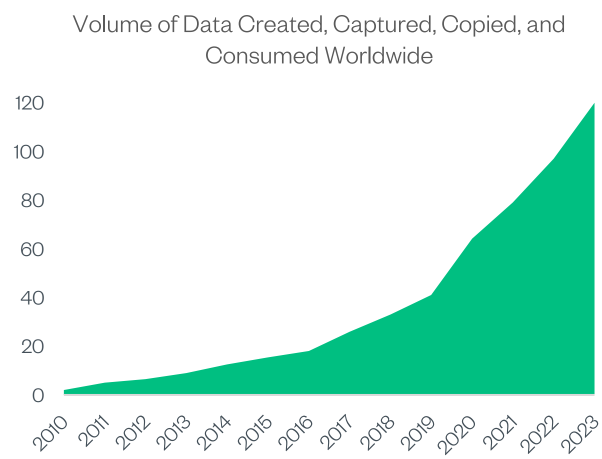


Speed of adoption is perhaps the most straightforward of the three components. However, ironically, it's also likely the most challenging and costly to implement. To capitalize here, a firm needs to integrate new technologies before they hit the mainstream market. An example is the use of sentiment analysis about a decade and a half ago. The early adopters who took Tim Loughran and Bill McDonald's pioneering work and applied it to their own transcript databases earned significant profits.[9] There were no guidelines or precedents for how sentiment analysis could be utilized, making it a risky and complex endeavor. That is exactly what made it highly profitable until the approach was gradually arbitraged away, and corporate management began to adapt their communications based on sentiment analysis. The same principle applied to text summarization and information extraction from tables. While these techniques were differentiators when they first emerged, they are now commercially available and commonly used by most investment firms.[10] There is nothing wrong with a fast-follower strategy, but every company needs to assess if they want/can be the first to apply certain techniques in their niche. As we’ll share later in the paper, we’ve consistently aimed to position ourselves at the forefront, being among the first long-term investors to adopt nascent technologies into our workflow. While we may not outpace giants like Citadel, we thankfully aren’t competing with them. We strive to utilize these technologies in a way that isn't yet commercially available and continue to stay ahead of the curve.

The second element of embracing AI revolves around a company’s unique and idiosyncratic data. If there is a durable competitive advantage with AI, it will be rooted in data. If the firm’s data sources, libraries, and tools are identical to those of others, how can they generate a unique insight? Perhaps

[9] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1331573
[10] These techniques are now available from Capital IQ, FactSet, and AlphaSense.

they'll hire the Magnus Carlsen of data science, but we deem this unlikely.[11] Good data scientists know there is huge potential in the vast amount of data that companies already generate but fail to capture. Consider routine calls with management teams: are they recorded, transcribed, and logged for future reference? Are unique terms and phrases from those calls fed back into your processes? What about weekly research meetings and the decisions made therein? Is there a record of who made what decisions based on what data? Internal research reports offer another goldmine. Are they produced in a machine-readable format that can feed AI models with insights and conclusions? The journey from zero to high-quality data may be long and daunting, but many companies haven't even begun to capture the simple data they generate. Do they know how many research reports every analyst generates each quarter? If you have a unique aspect in your process, capturing this data is the first step to magnifying it over time with AI. We'll share more of our approach to data throughout this primer, but our modus operandi is to capture as much data and metadata as possible. We believe that even if it isn't useful now, it may be later.



Graph of data volume over the past 13 years in Zettabytes. 1 ZB = 1 Billion TB
Data sourced from Statista

The final component of embracing AI centers on reshaping a firm's workflow to optimally use these new technologies. Allow us to explain. The Greeks are often quoted as having said, "To know a thing is to do a thing."[12] Most firms had a team member read Daniel Kahneman's "Thinking, Fast and Slow," likely the most referenced book of the last decade. In it Kahneman explored how decision-making environments could be designed to engage our slower analytical brains and minimize our reactive instinctual brains. Firms also read Ray Dalio's "Principles: Life and Work," where he discusses structuring trades based on a comprehensive understanding of economic cycles, ultimately coding these ideas into a machine for automated execution. James Montier and the entire field of behavioral investing are also salient. Montier emphasized how our cognitive biases lead to subpar investment decisions and outlines specific actions that can be taken to counteract these biases. Every firm knows the ideas of Kahneman, Dalio, and behavioral investing, but how many have integrated them into their process? If a firm hasn't integrated these accepted ideas into practice, how will they integrate more esoteric technological innovation? Black-lining, sentiment analysis, and simple database APIs are now widely available. Are they being utilized? The most crucial element of embracing AI is not learning about new ideas and technologies but reshaping your workflow to optimally use them. The goal is to increase the efficiency, without hurting the quality, of our most valuable asset – our human analysts.

[11] Magnus Carlsen is the highest-rated Chess Grandmaster of all time.
[12] While this exact quote does not appear in ancient Greek literature, the idea is contained in this phrase: "Τά γὰρ ἐνδεχόμενα μαθεῖν πρότερον ἢ κατὰ τὴν ἐνέργειαν γίνεται, ταῦτα μαθόντες οὐ ποιοῦμεν." (For the things we have to learn before we can do them, we learn by doing them) – Aristotle, Nichomachean Ethics

At Saguaro, we embrace AI by visualizing research and data science as two separate circles, gradually moving toward convergence. Over time, it will be increasingly difficult to differentiate between our research and data science efforts. Although many have traditionally employed data science to support their research, our aim has been to reciprocate this relationship, asking our research team to also support our data science endeavors. This means our full team is committed to all three components of embracing AI. We speed adoption by intentionally training and using new technologies that the data science team brings to bear. We are committed to both capturing data in the day-to-day and to putting in the extra work to make our reports machine-readable. Finally, we view our process as evolutionary where it can consistently improve and grow as we find better ways tomorrow to do what we did today. Our ultimate goal in "using AI" is to evaluate and value every publicly traded company daily, and to do so with ever-improving quality.

"What you observe is of little or no import, unless it influences action."

How did you get into AI and how can we?

"AI?"
"You did what?!"
"You shared our process with IBM?"

The doors were open, and Bermuda's tropical breeze brushed the skin, but no island magic could relax this room's tension. My boss craned his neck in pained frustration but calmed his voice. "Start again. From the beginning."

It had been a few months before, in 2015. I was in my office reading a 13D Research newsletter discussing the looming AI disruption. Gazing out the window at Alabama's lush green canopy, the thoughts came rushing in: 1) if a computer can distinguish between pictures of a dog and cat, surely it could distinguish a great business from a terrible one, 2) some multi-billion dollar shop is going to make a lot of money with this, 3) this could eventually put us out of business, 4) wait… we are a multi-billion dollar shop, 5) why couldn't we do this? I stood still. Then walked over to the phone and called our Head of IT, Kevin Overlaur.

Photo of IBM Watson Competing on Jeopardy
Photo illustration sourced from IBM

A week later, Kevin and I had our introductory call with IBM. We got right to the point. "We'd like to deploy Watson for our use case."

"Okay, tell us about your team, how many programmers, developers, and data scientists do you have? What types of projects have you already deployed…"

"Um… before I answer that could you explain what a developer is?"

IBM didn't think we were ready for Watson. We pushed harder. "How about we connect you with a consultant partner who could build this on the same software as Watson?" Done. We agreed.

We explained to this new partner, Revelwood, and our lead consultant, Justin, exactly what we were looking to accomplish:[13] "We want to build a machine that can distinguish great businesses from mediocre and terrible ones. The hardest part of our process is finding great businesses. We can do the rest." Justin told us what it would take for a proof-of-concept (PoC): Data on 100 companies and $10K. This brings us back to Bermuda.

We were on the island to meet with a few re-insurance companies but decided to add-on a strategic retreat. It didn't feel add-on anymore. After discussing some of our automation efforts, Bruce Donnellan made a comment, "You know, this is great, but what we really ought to be doing is seeing if we could use AI to help us find great businesses." That's when I shared about IBM, Revelwood, and our

[13] https://revelwood.com/

Top-Secret project. A lengthy, heated, but robust discussion ensued. By the end, the insurance flavor of our trip won out. As "insurance" against future disruption from AI, the project would proceed.

Armed with sufficient cash and a team of summer interns (including one Bryan Tsiao), we set out to gather and clean ten years of both financial and textual data on 100 companies that we had already evaluated. Here, to the best of our memory, and excluding any proprietary information, is what we did.

We selected:

50 High-Quality Companies

50 Low-Quality Companies

For each company, we did the following:

- Pulled ten years of financial data: Income Statement, Balance Sheet, Cash Flow Statement.
- Ensured each company was "standard." No banks, insurers, REITs, etc.
- Pulled ten years of 10-Qs, 10-Ks, and earnings call transcripts.
- Transformed all data into a standardized machine-readable format.

If a company didn't have ten years of data, we didn't include it. This small proof-of-concept required roughly 9,000 text files and 100,000 financial data points. This simple sounding process took roughly 240 work-hours to complete as the data had to be standardized and error-free. When most people think of data-science or AI, they think of wizards playing with super-computers and cutting-edge algorithms. This couldn't be further from the truth: 95% of the work is blankly staring at a computer screen "cleaning" data.

Our software base was IBM's SPSS Modeler (not something we'd use today), a versatile piece of data-science software with a graphical user interface that allows non-programmers to manipulate data and deploy models for both text and numbers.[14] This is what it looks like:

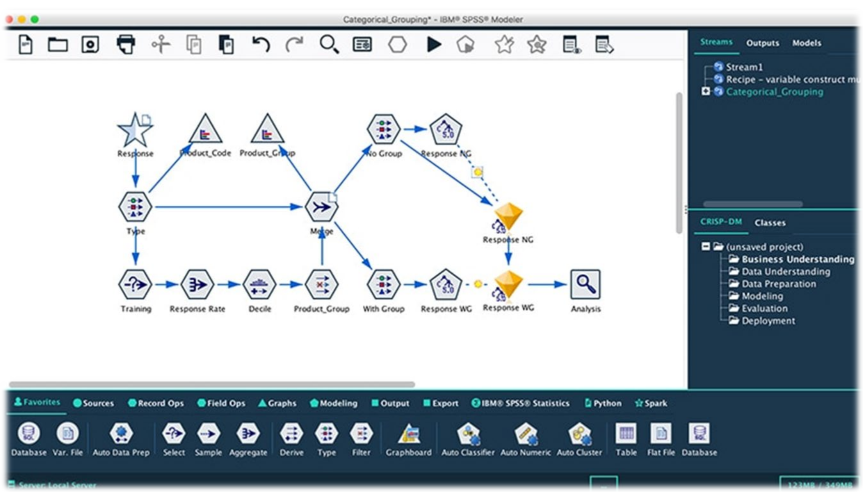


Image of IBM's SPSS Modeler for illustrative purposes only
Image sourced from IBM

[14] https://www.ibm.com/products/spss-modeler

With the data prepared (cleaned) and a system in place, we divided the companies into two groups: a training set of 70 and a testing set of 30 (please see the Breakdown of the 100 Companies graph on the next page). Using both the quantitative and textual data available for the 70 training companies, the machine attempted to create a "predictive" model that could distinguish good companies from bad. For this proof-of-concept, we didn't create features or refine the data in any way (for an explanation of feature engineering, please see the box below). SPSS Modeler has built-in text and numerical finance features which we used off-the-shelf. Given our low-performance computing available at the time, producing the model itself took about 40 hours of computer processing time.

Once the model was finished, we tested it on the thirty companies we'd held back. We provided the machine with the qualitative and quantitative data for every business and asked it to categorize each company as high-quality or low-quality. It was 80% accurate. So, we tried training different models from scratch using different breakdowns of the data (we put different companies in the group of either 70 or 30). This process is called cross-validation. The results came back similar no matter how we parsed the data. We were on to something, and we needed a bigger test. Obviously, the cost would go up (IBM doesn't give their software away, and you only get one free proof-of-concept), but the bigger question was how were we going to get all the data?[15]

Our second proof-of-concept aimed to generate a list of potentially qualifying companies, with the caveat that we'd never previously evaluated them. To find 10 to 20 promising names, the computer would need a lot of names to consider. We gave it 500. Our training-set remained the 100 companies we'd used in PoC #1. How we gathered that much data is a story in itself, which we'd be happy to share, and it took a long time. Ultimately, it led us down the first steps of automation, and we've never looked back. With the data in hand, we set the model loose and generated a list of potential companies that might qualify for our investment process. When the human research team examined the list suggested by the machine, they

Feature Engineering

Raw data and features have distinct meanings:

1. Raw data – refers to the unprocessed, unrefined information collected from various sources.
2. Features – refers to the processed, calculated, and refined information extracted from the raw data.

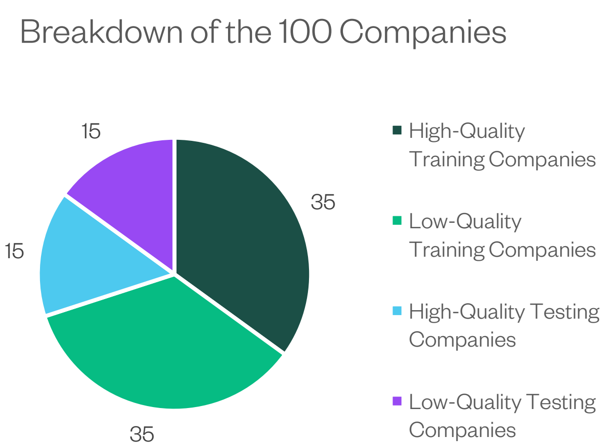
For example, let's say you're trying to predict housing prices based on real estate data. The raw data might include information like the address, year built, square footage, last sale price, number of bedrooms, and neighborhood demographics. From this raw data, you could create features such as:

- Age of the house (current year – year built)
- Price per square foot (total price/square foot)
- Average income in the neighborhood (extracted from the neighborhood demographics)

These features are derived from the raw data, but they represent specific characteristics that are more meaningful and useful for predicting house prices. By performing feature engineering, you can turn the raw data into a set of features that better capture the relationships and patterns needed for accurate predictions in a machine learning model.

[15] It may be hard to believe today, but in 2016 the data aggregators (CapIQ, Reuters, FactSet, etc.) did not offer data feeds for most of their products. Most quant firms of the time focused on pricing data and order books offered by exchanges, but even those who used qualitative data needed faster access than the aggregators could provide. Dow Jones attempted parsing their news wire feeds in the early 10s, but with little success. RavenPack is now the industry standard. Out of dumb-luck and sheer stupidity, we did manage to negotiate a deal for one provider to offer us a copy of their full database for a cool $1mn if we would provide the hard drives. Given the proof-of-concept nature, we said no as we believed the price too high.

found that 60% met or exceeded the investment quality bar. An astounding result, considering that the historical human-only hit rate was around 20%.



The next step was the real goal. Could we use this tech to identify non-US companies? Specifically, could this tech help us source enough ideas to make us truly global? Proof of Concept #3 was a test run on both the full Canadian and UK market. We had to scrape and clean all the data necessary for these two markets, and this data is significantly different from the US. We also decided to increase our training-set from the 100 companies mentioned above to all of the firm's historical data. In short, this was a major undertaking and given that we were trying to build a system/process that was repeatable and not just a one-time lift, it consumed the better part of a year. Cleaning data on thousands of companies is time-consuming, and the data of 5 years ago was not as standardized as it is today. Ultimately, the hit rate for both the UK and Canada was north of 50% and led to the greenlighting of a full global build-out.

This led to the creation of an AI-team, an end-to-end system built from scratch on proprietary code, and the exploration of applying these technologies to other parts of the research process. As these were proprietary to our previous firm and some are still in use today, they aren't ours to discuss. Also, everything mentioned above was performed a lifetime ago in the machine-learning world. Every piece of software and every line of code would be different today. Gathering the necessary data would be a different process as most of it is available via APIs. In addition, the entire process was dependent upon the proprietary data of our previous firm; neither we nor anyone else would be capable of re-building that system. It remains proprietary. As we've stated, AI is an amplifier and not a differentiator. What set us apart were the processes that generated the data in the first place, and the decision to embrace AI. That same decision is leading in new directions today.

The key takeaways for anyone interested in taking their own AI journey are straight-forward. First, you don't have to be an expert to get started, find a partner. Second, even with resources, this journey takes time, so get going. Finally, AI isn't a silver bullet. It can merely amplify the secret sauce you already possess. Only you know what that is.

"Did we mention that we globally brought down Capital IQ not once, but twice?"

Can you give long-term investment examples?

Image created by Midjourney using the description *"investing using AI inspired by the Apollo Program"*

Yes. As mentioned above, our goal in "using AI" is to analyze and value every publicly traded company daily and to do so with ever-improving quality. This is our tiny replica of the Apollo Program, if you will.[16] And like the Apollo Program, we've broken our large goal down into smaller more manageable components. Specifically, we've looked at our full research process from A-Z and decomposed it into the smallest possible steps. For each of those steps, we consistently ask if there is a technique or a technology we could use to automate, improve, or speed it up. As you can imagine there is low-hanging fruit, which we've pursued, and there are other elements that either don't look plausible or cost-efficient at this point. Simultaneously, the Apollo program was responsible for many unanticipated innovations that improved everyday life.[17] While none of our unanticipated innovations have led to commercialized products, they have improved our process in unexpected ways by adding steps that improve the quality of our work at very little additional cost. Below we will discuss a select few examples, including components of our research process where we use AI, an attempt where AI failed to improve performance, and where AI gave us a unique innovation we didn't previously have.

Company Categorization based on Quality = Idea Generation

The previous section described our initial steps into company categorization. We've never stopped improving these ideas and were fortunate when starting Saguaro to be able to build a similar system again from the ground up with a clean sheet of paper. This has led to a plethora of new data sources, a much more robust process to capture internally generated data, and a more modular format that allows us to take virtually any other tool we develop and add it to this system to see if it adds value in identifying high-quality businesses. While we don't want to share the specifics of what we do at every step for obvious competitive reasons, we do provide a general outline of this system in the section on the specific tools we use. The output of this system couldn't be simpler: it's just an Excel file that our analysts can open. It lists companies we've never analyzed before that the machine believes have the highest probability of exceeding our quality bar. Here's an example:

[16] The goal of the Apollo program was to land a man on the moon and return him safely to Earth. The actual work, however, could only be done on the sub-components of this goal: lunar lander, guidance system, rocket engines, etc.
[17] Examples include solar panels, quartz clocks, thermal blankets, Velcro, the Dustbuster, vacuum-sealed food, zero-gravity pens, large-scale shock absorbers, and more.

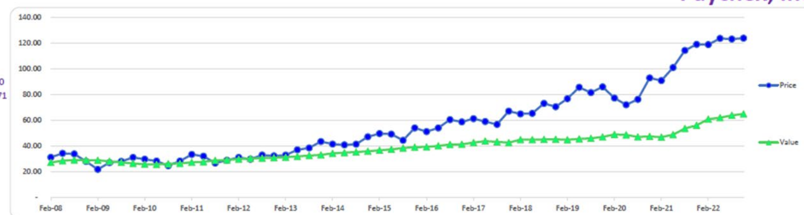
	A	B	C	D	E
1	scm_id	company_name	tickersymbol	exchangesymbol	score
2	22178	Paychex, Inc.	PAYX	NASDAQGS	0.895
3	73888	Novo Nordisk A/S	NOVO B	CPSE	0.879
4	22350	ResMed Inc.	RMD	NYSE	0.870
5	23040	Edwards Lifesciences Corporation	EW	NYSE	0.854
6	60231	Public Storage	PSA	NYSE	0.849
7	21057	Mettler-Toledo International Inc.	MTD	NYSE	0.848
8	72418	Zoetis Inc.	ZTS	NYSE	0.844
9	52069	Evolution AB (publ)	EVO	OM	0.837
10	38984	TransDigm Group Incorporated	TDG	NYSE	0.832
11	26881	Intercontinental Exchange, Inc.	ICE	NYSE	0.829
12	21979	KLA Corporation	KLAC	NASDAQGS	0.825
13	21345	The Hershey Company	HSY	NYSE	0.820
14	23641	International Container Terminal Services, Inc.	ICT	PSE	0.818

Image of an Excel output generated using AI for illustrative purposes only
Photo illustration sourced from internal Saguaro data

Output can be customized based on market cap, geography, or even the number of Wall Street analysts covering a given name. The obvious follow-up question is how this differs from a stock screener. First, most screeners don't incorporate textual data. We train our system to identify both competitive advantages and risks that appear in company filings. Second, when using a screener, if you put in more than 6 or 7 variables, you are likely to get zero results. We currently utilize the 138 most predictive features (variables) from the thousands we've evaluated and are constantly looking to expand this list. No matter how many variables we use, we will still get a full list of results. In essence, machine learning tools allow us to build a machine that analyzes companies in a way similar to human research analysts. While our human analysts can better analyze any individual company than the machine, the machine can analyze every public company in the world in fifteen minutes. This equals idea generation. To aid our human analysts when they receive the above list, the system also creates a one-page overview of each company. This is merely automation supplementing the AI:

Ticker:	NASDAQGS:PAYX	Date:	3/23/2023
CIQ Full Ticker:	NasdaqGS:PAYX	LFR FY:	2022
CIQ ID:	IQ295368	Fiscal Year End:	5/31
SEC CIK Code:	CIK_0000723531	Currency:	USD
ISIN:	I_US7043261079	Exchange:	NasdaqGS
State of Inc:	Delaware	ISO Code:	US
Country of Inc:	United States	# of Employees:	16,000
Est Earnings:	3/29/2023	Year Founded:	1971
Conf Earnings:	3/29/2023	WS Analysts:	39
Parent Co:	Paychex, Inc.	Short Interest:	2.59%
Parent Ticker:	NasdaqGS:PAYX	Insiders Own:	10.61%
Website:	www.paychex.com		
Country:	United States		

Paychex, Inc.



Address:	911 Panorama Trail South, Rochester, New York, 14625-2396, United States
Industry Sector:	Industrials
Industry Group:	Commercial and Professional Services
Industry:	Professional Services
Prim Industry:	Human Resource and Employment Services
SIC Code:	8721
Company Type:	Public Company
Status:	Operating
Template Type:	Standard

USD Mkt Cap:	39,007
Native Mkt Cap:	39,007
USD Net Debt:	(387)
Native Net Debt:	(387)
10yr FCF/NI =	110.8%
Avg Dbt/EBITDA =	(0.2)

Notes:

Basic Numbers	2013	2014	2015	2016	2017	2018	2019	2020	2021	202
y/y rev growth		*8.3%*	*8.8%*	*7.7%*	*6.8%*	*7.1%*	*11.7%*	*7.1%*	*0.4%*	*13.7%*
Revenue	2,326	2,519	2,740	2,952	3,153	3,378	3,773	4,041	4,057	4,612
Adj. EBIT	905	983	1,054	1,147	1,254	1,292	1,377	1,461	1,461	1,840
Net Income	569	628	675	757	826	994	1,034	1,098	1,098	1,393
Levered FCF	577	797	792	921	866	1,122	1,148	1,314	1,146	1,373
Margins										
Adj. EBIT	38.9%	39.0%	38.5%	38.8%	39.8%	38.2%	36.5%	36.1%	36.0%	39.9%
Net Income	24.5%	24.9%	24.6%	25.6%	26.2%	29.4%	27.4%	27.2%	27.1%	30.2%
Levered FCF	24.8%	31.6%	28.9%	31.2%	27.5%	33.2%	30.4%	32.5%	28.2%	29.8%
CapEx	4.2%	3.3%	3.8%	3.3%	3.0%	4.6%	3.3%	3.1%	2.8%	2.9%
Key Metrics										
Adj. EBIT/Tangible Assets	83.4%	86.8%	93.2%	106.3%	120.8%	93.6%	47.6%	47.8%	46.7%	56.1%
Net Income/Common Equity	32.1%	35.3%	37.8%	39.6%	42.3%	42.2%	39.5%	39.5%	37.2%	45.1%
FCF/NI	101.3%	127.0%	117.4%	121.6%	104.8%	112.9%	111.0%	119.7%	104.4%	98.6%
Other Metrics										
Net Debt (Cash)/EBITDA	(0.5)	(0.5)	(0.5)	(0.3)	(0.2)	(0.3)	0.1	0.0	(0.1)	(0.2
Diluted Shares	365	366	365	363	363	362	362	361	362	363
% of FCF										
Debt Repayment (Issue)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-69.4%	-0.4%	-0.2%	-0.1%
Net Share Repurchase (Issue)	-12.6%	17.1%	16.9%	8.9%	15.9%	12.7%	2.4%	11.5%	6.1%	8.9%
Dividends	82.7%	64.1%	69.6%	65.9%	76.5%	65.9%	72.0%	67.7%	79.3%	72.8%
Cash Acquisitions	3.7%	1.2%	3.4%	32.2%	0.0%	16.1%	86.5%	0.5%	1.7%	1.8%
Divestitures	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Paychex, Inc. provides integrated human capital management solutions for human resources (HR), payroll, benefits, and insurance services for small to medium-sized businesses in the United States, Europe, and India. It offers payroll processing services; payroll tax administration services; employee payment services; and regulatory compliance services, such as new-hire reporting and garnishment processing. The company also provides HR solutions, including payroll, employer compliance, HR and employee benefits administration, risk management outsourcing, and the on-site availability of a professionally trained HR representative; and retirement services administration, including plan implementation, ongoing compliance with government regulations, employee and employer reporting, participant and employer online access, electronic funds transfer, and other administrative services. In addition, it offers cloud-based HR administration software products for employee benefits management and administration, time and attendance, digital communication solutions, recruiting, and onboarding solutions; plan administration outsourcing and state unemployment insurance services; various business services to small to medium-sized businesses comprising payroll funding and outsourcing services, which include payroll processing, invoicing, and tax preparation; and payment processing services, financial fitness programs, and a small-business loan resource center. Further, the company provides insurance services for property and casualty coverage, such as workers' compensation, business-owner policies, cyber security protection, and commercial auto, as well as health and benefits coverage, including health, dental, vision, and life. It markets and sells its services primarily through its direct sales force. The company was founded in 1971 and is headquartered in Rochester, New York.

Management: Name:	Title:	Total Comp (mm):	Stake %:	USD Stake:
Gibson, John	President, CEO & Director	2.51	0.01%	5.65
Rivera, Efrain	Senior VP & CFO	2.87	0.02%	8.49
Bottini, Mark	Senior Vice President of Sales	2.52	0.02%	8.12
Gioja, Michael	Senior Vice President of Product Develo	2.52	0.01%	2.48
Schaufenbuel, Bradley	VP & Chief Information Security Officer	-	No Data	No Data
Schrader, Robert	Vice President of Finance & Investors R	-	0.00%	0.98
Schaeffer, Stephanie	VP, Chief Legal & Ethics Officer and Cor	-	0.02%	6.97
Fiorille, Frank	Vice President of Risk, Compliance & Da	-	No Data	No Data
Collins, Neal	VP of Corporate Development and MD	-	No Data	No Data

Board: Name:	Title:	Total Comp (mm):	Stake %:	USD Stake
Mucci, Martin	Chairman	9.63	10.18%	39.30
Gibson, John	President, CEO & Director	2.51	1.47%	5.65
Flaschen, David	Independent Director	-	1.36%	5.26
Tucci, Joseph	Lead Independent Director	-	1.46%	5.64
Joseph, Pamela	Independent Director	-	0.24%	0.92
Velli, Joseph	Independent Director			
Golisano, B.	Founder & Director			
Doody, Joseph	Independent Director			
Bonadio, Thomas	Independent Director			

SAGUARO CAPITAL MANAGEMENT

Image of an automated Excel output for illustrative purposes only
Photo illustration sourced from internal Saguaro data

Combined, the list and the one-pagers have been a game-changer for our idea generation. The list we provided above includes many names that professional analysts would be familiar with. Any company we have not fully reviewed is included, even if we already know that it is a strong business. The system has been most helpful in the Smid cap arena where many great businesses are still emerging. It adds the most value in three specific circumstances: 1) Evaluating niches where we have less familiarity: National Research Corporation – NRC, 2) Exploring geographies that would be further down our list: Mainfreight Ltd – NZE:MFT, and 3) Forcing consideration of companies we are naturally biased against: Games Workshop Group PLC – LON:GAW. We are happy to go into greater depth, provide a live demo, or even go through the code in a non-public setting. And we love talking companies, so please ask!

Summarization of Financial Documents = Data Gathering and Decision Making

One of the key applications of large language models (LLMs) is the summarization of lengthy financial documents. Human analysts spend a lot of time reading. Anything that can reduce this load or help us better prioritize is a winner. For instance, if you didn't want to read four pages about our journey into AI, here is GPT 4's summary of that section:

> The passage is a first-person account of the author's journey in leveraging artificial intelligence (AI) in their multi-billion-dollar business, focusing on integrating AI to better distinguish great businesses from mediocre or poor ones. Initially unfamiliar with the technology, the author began this journey after reading a newsletter about the impending AI disruption in 2015. They partnered with IBM and consulting firm Revelwood to deploy IBM's Watson for their specific needs. Despite the tension and skepticism within their team, they proceeded with a proof-of-concept (PoC), gathering and cleaning financial and textual data of 100 companies. Using IBM's SPSS Modeler software, the author trained the AI to differentiate between good and bad businesses, achieving 80% accuracy. After further tests and another PoC involving 500 companies, the human research team found 60% of the machine-selected businesses met or exceeded the investment quality bar, significantly higher than their historical 20% success rate. The author further tested the technology on Canadian and UK markets, leading to the creation of an AI team and the development of a global investment strategy. The author highlights the rapid evolution of AI technology and emphasizes the importance of partnerships and self-assessment in utilizing AI.

Image of ChatGPT's summary of the Saguaro AI Primer
Photo illustration sourced from OpenAI

Models such as Bloomberg GPT have been trained on a large corpus of financial documents and can condense complex financial text into more manageable insights.[18] Unfortunately, today most large language models can only summarize roughly 2,000-3,000 words before hitting their limit.[19] However, newer models are being developed to overcome this obstacle.[20] Humorously, we once spent nearly twelve months trying to develop text summarization capabilities prior to large language models. We only succeeded in producing gobbledygook. This was true especially when financial or numerical data was interspersed with the text. Early models had no ability to understand data in that format and therefore couldn't successfully parse the numbers.

As the size limitations on summarization are removed, these tools will become increasingly valuable. Our short-term plan is to use large language models to summarize our previous discussions, reports, and viewpoints on individual companies. This will help us remember what has previously been done and make it faster to find specific information. The obvious derivative from these models is the ability to answer specific questions or queries from the text. While the accuracy for concepts and specific quotes is high, models are not yet at the point where they can accurately summarize numerical conclusions. We'll be excited to demonstrate this model once available.

[18] https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/
[19] Chat GPT 4.0 for instance has a token limit typically around 4,096 which equates to about 2,000-3,000 words.
[20] https://www.mosaicml.com/blog/mpt-7b

Predicting Revenue Growth = Valuation

For some, investing = valuation, and their question is if AI can help here. It will, but with limits. We have tried to predict 5-year company growth rates using historic data. We failed. No matter how many times or ways we've attempted to build a model that can predict long-term revenue growth, we have never achieved a result better than random chance.[21] This undertaking is much more challenging than you'd imagine. Do you want to calculate organic or non-organic growth? How do you parse management growth predictions from the text where available? Are they referring to top-line or bottom-line growth? Is the growth for the whole company or just for a segment? Is it a three-year prediction or a five-year prediction? Can you pull information from charts and graphs?

We've tried every type of model. Fail. We gathered even more data. Fail. Finally, we even gathered every human prediction we could find from any source. They were all no better than random chance. While this made the data science team feel better, it also led to our current thesis that the future is unknowable and that even AI can't predict the unknowable. It is possible that other teams have solved this problem, but we find it unlikely. Still, this effort wasn't for naught. We learned, created new tools, and shifted our focus to things we can predict. It also radically strengthened our conviction that thinking of valuation in terms of ranges is superior to point-estimates of value given investing's inherent uncertainties. We have no plans to attempt using AI to predict future revenue growth again.[22]

Text Classification = Data Gathering = Decision Making = Portfolio Management

The next step for our AI system involves leveraging the long tail of information available for individual businesses and industries. To achieve this, we must first establish which documents are relevant for which companies. While most documents come pre-labeled, many remain unlabeled, and some labeled documents may pertain to multiple businesses.[23] We are creating this system, which we will explain further later in the paper, as we can't handle the quantity of text documents available without it. This technology shares commonalities with the system used in spam email detection and document labeling. In addition to assigning documents to various companies and industries, we'd eventually also like to suggest potential competitive advantages and risks that may be mentioned in these documents. We aim to do this without introducing any bias. Moreover, this technology could prove useful in pinpointing documents relevant to the ongoing topics of discussion within the research team each week. One other notable subfield of text classification that we've used is sentiment analysis.

We've addressed sentiment analysis in other sections of this primer and our approach has been fairly industry standard. However, we haven't found this technique to be predictive for company quality and it doesn't typically factor into our long-term thesis. We ultimately plan to use sentiment analysis to help understand why specific names may be discounted at any particular time by rating overall management, analyst, and broad market sentiment. Any time we choose to buy or sell a security, we

---

[21] For this exercise we defined random chance as correctly predicting if company growth was above or below the average growth rate for all companies, and if so by how much.
[22] With the caveat that a landmark paper or technological break-through could change our thinking. Nevertheless we currently believe that accurately predicting the future will always belong to the realm of fantasy and not AI.
[23] A Label here just means that a document or information is associated with the proper company(ies).

take a divergent view. Understanding management, analyst, and market sentiment will give us further insight into where this divergence is based. Given our philosophy and process, building this capability is lower on our list of priorities, but is well-established and now offered in many commercially available products.

Implied Probability Distribution of Stock Price = Valuation = Portfolio Management

A side benefit of developing comfort with large quantities of data is using these tools to extract information from the market itself. We do this in a number of ways, but one of the most interesting is using the options market to understand Wall Street's assumptions about where a stock's price is likely to go over the next year or two. As stated above, when we purchase a stock, we are taking a divergent view from the market. If everyone thought the company was worth more, it wouldn't be discounted. To ensure that our ranges make sense, we triangulate numerous ways, including an implied probability distribution of a company's stock price one to two years in the future. We generate this insight using a company's options data.

The first time we heard someone mention a probability distribution for stock prices our eyes glazed over, but we quickly realized that understanding what the market is thinking about the future for any stock is incredibly valuable. For instance, if the market is very confident in the future price of a stock, the distribution or range of outcomes for the stock is going to be tight or narrow. If the market is uncertain, the outcome will be wide. Also, distributions start to look abnormal if bankruptcy or a takeover is a possibility. Building this crystal ball that reveals the market's deepest secrets just takes access to a company's options data. Options are priced at many different strike prices for several expiry dates in the future. The cost of those options at each strike price provides enough data to construct the curve we are talking about. This process requires use of the Black-Scholes model and the construction of thousands of theoretical option prices. The exact methodology of this process is beyond the scope of this document, but Morgan Stanley has concisely summarized it.[24] The result is demonstrated below.

[24] https://www.morganstanley.com/content/dam/msdotcom/en/assets/pdfs/Options_Probabilities_Exhibit_Link.pdf

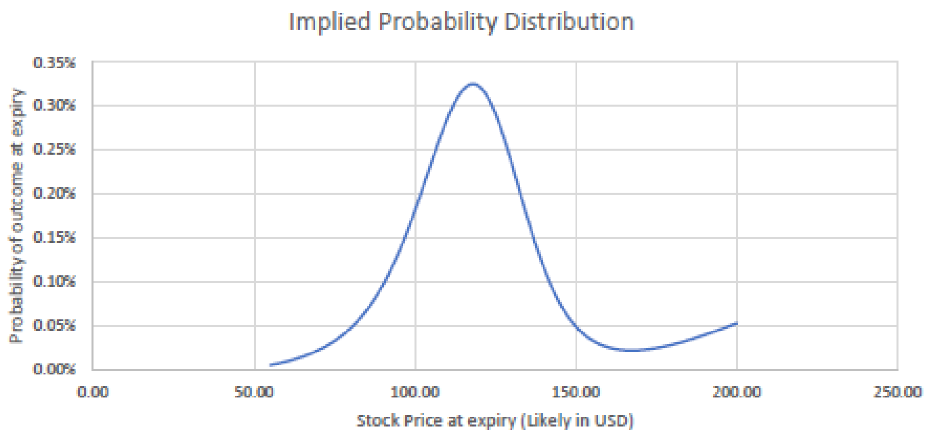
PAYCHEX INC

Inputs	
Stock Symbol (US-only):	PAYX
Option Chain Expiry Date:	1/19/2024
Risk-Free Rate Assumption:	4.25%
Today's Date:	6/12/2023

Results	
Probability of close below lowest available strike:	-1.73%
Projected Stock Price at Expiry:	118.07
Current Stock Price:	113.09
Days until Expiry:	221.00
Expected Return above Risk-Free Rate:	7.38%

Results for export to the company Range worksheet:

Min(1)	Low(2)	High(3)	Max(4)
87.48	94.44	147.95	172.45



How to Interpret:
An "implied probability distribution" shows us the market's current expectation for the future price of any stock at a specific date in the future. Obviously, that future date is the Option expiry date we enter above. A narrow range shows us that the market has a lot of confidence around a particular price, and wider ranges demonstrate lower confidence. Companies with higher debt, higher growth, or just more uncertainty will have wider ranges than their counter-parts. As you might expect, the further the date in the future, the wider the range should be. You'll also notice strange shapes if companies face bankruptcy risk or a high probability of being acquired.

Methodology:
We took inspiration from Morgan Stanley's Implied Probability offering. We use all available options data at a given expiry to construct an implied volatility curve. This is done using the Black-Scholes model. The implied volatility curve is then used to generate 1,000 theorhetical call option prices. Finally, we construct butterfly spreads at each of these 1,000 price points, and calculate the probability of the stock closing within that spread at expiry by taking the cost of the spread divided by the maximum payoff.

Notes:
1.) 95% of expected prices are higher (Bottom-end of 90% confidence interval)
2.) 90% of expected prices are higher (Bottom-end of 80% confidence interval)
3.) 90% of expected prices are below (High-end of 80% confidence interval)
4.) 95% of expected prices are below (High-end of 80% confidence interval)

SAGUARO
CAPITAL MANAGEMENT

Image of a probability distribution for Paychex's stock price for illustrative purposes only
Photo illustration sourced from internal Saguaro data

A quick glance at this output for any name we are valuing is a helpful check for our thinking. The same is true whenever we are considering entering or exiting a position. We'd be happy to demonstrate this system at any time.

Per Share Intrinsic Value Growth = Idea Generation = Decision Making

Of the unanticipated innovations provided to us by our AI moonshot, this is one of our favorites. For many years, even prior to working with AI, we'd been searching for one metric to rule them all when it came to expressing quality. During a particularly hectic season, while crafting machine-learning features, and analyzing businesses, we reflected upon a saying learned from Charlie Munger: "the best predictor of future behavior is past behavior." We want companies that will compound their intrinsic value per share for many years to come. The best predictor of future compounding is past compounding. While stock prices follow intrinsic value over the long run, they can be so erratic in the short run that they hold little predictive value.

What we needed was a way to show actual intrinsic value growth for every company on the planet over a full market cycle, say the last 10-15 years. The problems were legion: valuation takes time, even our best valuation is wrong, and as we stated above, computers can't develop reasonable growth estimates. The solution came from a debate on a particularly high growth business. Despite disagreement on how to value the company, we realized that if we held our valuation methodology constant from quarter to quarter, we could see that the company's value per share grew quarter over quarter. It hit me. What we needed to determine quality wasn't the actual valuation per share, but rather just the growth of that valuation per share over time. We could be wrong, but as long as we were consistently wrong, we could use the machine to extract the growth in intrinsic value per share for every company on Earth.

Revisiting Paychex, let's look more closely at its chart:

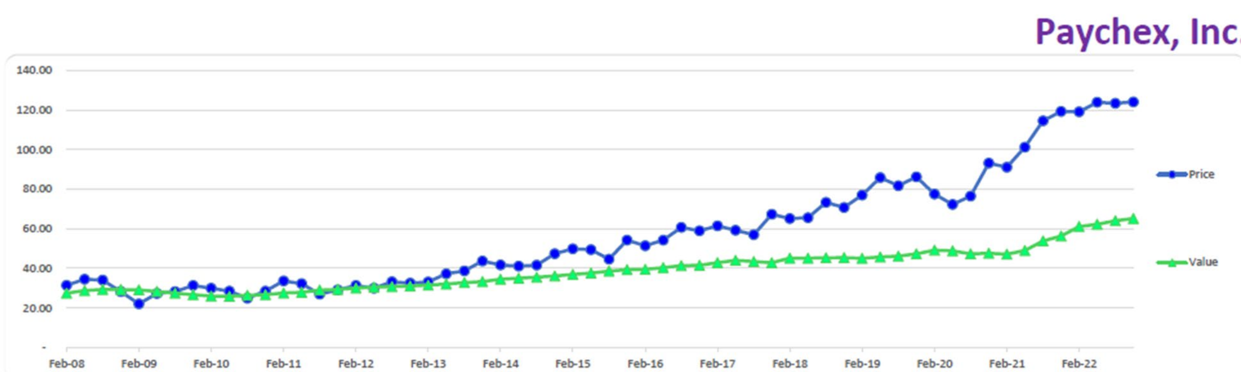


Image of a Paychex's intrinsic value and price chart for illustrative purposes only
Photo illustration sourced from internal Saguaro data

The green line above represents what we have been discussing. While this estimate of intrinsic value per share is wrong at every single individual point, the shape of the overall line accurately represents the growth or decline of intrinsic value per share on a quarterly basis over the full 15 years displayed. As it is a per share value, it includes all capital allocation decisions made by the various management teams whether acquisitions, divestitures, share issuances, or buybacks, etc.[25] While a stable green line

[25] Dividends are not reflected and need to be accounted for separately.

that is up and to the right doesn't guarantee a high-quality business, we've yet to find another metric that provides as much insight with a single glance.

In this section, we've provided six concrete examples of long-term investing embracing AI. Company categorization, text summarization, and text classification are specific instances of machine learning and natural language processing applied to improving an investment process. Predicting revenue growth was an example where reality fell short of the hype. The implied probability distribution of a stock's price is an example of using our data and the techniques we've developed to produce value in a non-AI way, though one that would not have occurred without the embrace of AI. Finally, the per share intrinsic value calculation demonstrated an unexpected innovation resulting from this process. We hope you take two things away from this discussion: 1) The runway for this type of innovation is incredibly long for a firm embracing this revolution, and 2) There is no element, component, or step in the value investing process that will not be impacted by AI in the future.

"The best way to have a good idea is to have a lot of ideas."

What tech stack do you use?

Everything we use is widely available. Unfortunately, it all requires a certain level of technical knowledge and ability. Most investment professionals, including our own research team, do not possess this same knowledge and ability. Nevertheless, in the journey of applying AI and data science to investment management, these tools and technologies are essential. While this section provides an overview of potential resources for data collection, storage, preprocessing, processing, and output, it does not discuss how to develop your technical competence. For help there, please see the additional resources section of this primer. The tools listed below will change with time, but this is a brief introduction to our current view of the ecosystem.

Data Collection

Data is the foundation of any AI and data science venture. The choice of data collection tools depends largely on the type of data needed. Web scraping libraries like Beautiful Soup and Scrapy in Python are useful for extracting data from websites. For financial data, APIs and data feeds from S&P, FactSet, and Refinitiv are popular choices. For social media data, the Twitter API or Facebook's Graph API can be employed to gather large amounts of data.[26]

Beyond relying on traditional sources, we also actively pursue idiosyncratic data. The primary source of this idiosyncratic data is our own research process. In essence, our research and data science teams operate in a powerful data-collection flywheel. The research team creates data during their company analyses, which is then captured and utilized by our data science team. This data forms the backbone of our predictive models. As we describe throughout this report, these models constantly feed new ideas to our research team. And this inevitably leads the research team to generate ever more data. This flywheel process continuously refines, optimizes, and ensures that our models reflect our particular research process.

Data Storage and Databases

Once collected, data needs to be stored efficiently. Structured data is typically stored in SQL databases like PostgreSQL or Microsoft SQL Server, while NoSQL databases like MongoDB or Cassandra are better suited for unstructured data such as raw financial filings.[27]

---

[26] Beautiful Soup: code.launchpad.net/beautifulsoup/
Scrapy: https://scrapy.org/
S&P Capital IQ API: Here
FactSet: https://go.factset.com/marketplace/catalog/product/factset-ondemand-api
Refinitiv: https://www.refinitiv.com/en/financial-data/company-data/company-fundamentals-data
Twitter: https://developer.twitter.com/en/products/twitter-api
Facebook Graph: https://developers.facebook.com/docs/graph-api/
[27] PostgreSQL: https://www.postgresql.org/
Microsoft SQL Server: https://www.microsoft.com/en-us/sql-server
MongoDB: https://www.mongodb.com/
Cassandra: https://cassandra.apache.org/_/index.html

When storing large amounts of documents, model weights, database backups, and other large data objects, cloud-based solutions such as Amazon S3, Azure Blob, and Google Cloud Storage provide scalable and secure storage options.[28]

Data Processing

Data preprocessing involves cleaning and transforming raw data into a suitable format for analysis. 'Pandas/'Polars' and 'cuDF' are popular libraries for processing tabular data on CPUs or GPUs respectively. When tabular datasets get too large for a single computer, 'Apache Spark' and 'Dask' are popular frameworks for scaling workloads across multiple machines.[29]

When processing large amounts of text data, it is common to switch to higher-performance languages than Python to scale the workloads. High-performance libraries have been wrapped in Python, but the language's limitations make it difficult to scale workloads across all available compute. Some example high-performance languages are Julia, Rust, or C++, where it is trivial to scale large workloads.[30]

Predictive Modeling

For machine learning, Python is the most popular programming language due to its powerful libraries and community support. To achieve high-performance machine learning, almost every machine learning library is a Code Wrapper around another language.[31] For tabular machine learning, 'Scikit-learn' and 'XGBoost' are some of the most popular libraries. 'Jax' and 'PyTorch' are go-to libraries for deep learning applications.[32]

For natural language processing (NLP) tasks, anything from sentiment analysis to text generation, the 'Transformers' Python library from Hugging Face is a commonly used tool. This library supports

---

[28] Amazon S3: https://aws.amazon.com/s3/
Azure Blob: https://azure.microsoft.com/en-us/products/storage/blobs
Google Cloud Storage: https://cloud.google.com/storage
[29] Pandas: https://pandas.pydata.org/
Polars: https://www.pola.rs/
cuDF: https://github.com/rapidsai/cudf
Apache Spark: https://spark.apache.org/
Dask: https://www.dask.org/
[30] Python: https://www.python.org/
Julia: https://julialang.org/
Rust: https://www.rust-lang.org/
C++: https://isocpp.org/
[31] A code wrapper, in the context of programming languages, is a way to interact with a piece of code written in one language from within another language. The wrapper essentially "translates" between the two languages, enabling them to communicate with one another.
[32] Scikit-learn: https://scikit-learn.org/stable/
XGBoost: https://xgboost.readthedocs.io/en/stable/
Jax: https://github.com/google/jax
Pytorch: https://pytorch.org/

several deep learning backends, so it gives users the flexibility to use the libraries they are most comfortable with. Additionally, you can start playing with models with just a few lines of code, so this library has greatly lowered the barrier to accessing countless state-of-the-art models. Below we have attached a snippet of code that predicts the sentiment of a sentence using the 'Transformers' library. Training your own model, or efficiently deploying them requires a lot more code, but libraries like this have democratized access to a variety of models.[33]

```
>>> import transformers

>>> # Load a transformers pipeline model
>>> model = "SaguaroCapital/saguaro-bert-sentiment"
>>> pipe = transformers.pipeline("sentiment-analysis", model)

>>> # Predict the sentiment
>>> prediction = pipe(["NVIDIA grew operating revenue by 15% while its
competitors remained flat for the quarter."])
[{'label': 'POSITIVE', 'score': 0.9999507665634155}]
```

Image of code to demonstrate a transformer for illustrative purposes only
Photo illustration sourced from internal Saguaro data

For training several billion parameter language models from scratch, EleutherAI's 'gpt-neox' or MosaicML's 'LLM Foundry' are common starting places for 'PyTorch' models and Google's 'maxtext' or 'paxml' is the 'Jax' equivalent. This is not where we would recommend starting.[34]

Output and Visualization

Once the data has been processed and analyzed, visualization makes the results intelligible. Visualization can help to communicate the results of the analysis to others and can help to identify patterns and trends in the data. Python's Matplotlib and Seaborn are animated, and interactive visualizations. For interactive dashboards, Tableau, Microsoft Power BI, and open-source libraries like Dash for Python can be used. The graphs we've published in previous sections are examples of visualizations. Many Excel users tend not to be impressed by simple visualizations as they are core features in Excel. Frankly, exporting raw output data into Excel can lead to some very robust visualizations for internal use.[35]

---

[33] Transformers from Hugging Face: https://huggingface.co/docs/transformers/index
[34] GPT-neox: https://github.com/EleutherAI/gpt-neox
LLM Foundry: https://github.com/mosaicml/llm-foundry
Maxtext: https://github.com/google/maxtext
Paxml: https://github.com/google/paxml
[35] Matplotlib: https://matplotlib.org/
Seaborn: https://seaborn.pydata.org/
Tableau: https://www.tableau.com/
Microsoft Power BI: https://powerbi.microsoft.com/en-us/
Dash: https://plotly.com/dash/

Regardless of methodology, data visualization is a critical step in data analysis and model explainability (also referred to as "interpretability"). The most common method for explaining model outputs is SHAP (SHapley Additive exPlanations).[36] For tabular problems, you can see what factors had the largest impact on the final prediction, and for NLP models, you can generate heatmaps of the attention given to each word to see the sections of text that had the largest impact on the final prediction.

Below, we have included an example of SHAP being used on a model trained for sentiment analysis. The darker the red or green on a word, the larger the negative or positive impact respectively on the final sentiment prediction. This model was trained to predict the sentiment of financial news headlines. With this visualization, we can see which tokens had the largest impact on the final predictions. In this example, "grew" had the largest positive impact on the final prediction. How the model determined these weights is still a black box.

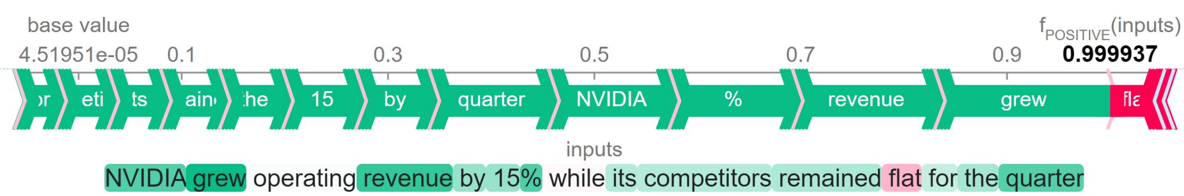


Image of sentiment prediction in action for illustrative purposes only
Photo illustration sourced from internal Saguaro data

Hardware

Machine-learning models have transformed the industry's approach to hardware. No longer is performance the sole criterion by which everything is judged. Today we must also consider power efficiency. High-performance CPUs and GPUs are a given for most machine-learning tasks. NVIDIA's CUDA-enabled GPUs are the most widely used for deep learning applications. A decade ago, the state-of-the-art computer vision model, AlexNet, was trained on two consumer-grade GPUs. This wouldn't fly today as modern models are commonly trained on hundreds or even thousands of data center GPUs or AI accelerators. The current best GPU is the H100 from NVIDIA, and servers usually come with 8 of them.[37] These servers are then connected and scaled to create clusters that allow for the training of large models that are too big to fit on a single machine. Alternatives such as Google's TPUs can be more powerful and cost-effective in certain scenarios, but they are significantly less popular, and they are only available on Google Cloud.[38]

[36] SHAP: https://shap.readthedocs.io/en/latest/
[37] NVIDIA DGX H100: https://www.nvidia.com/en-us/data-center/dgx-h100/
[38] NVIDIA's CUDA enabled GPU: https://developer.nvidia.com/cuda-gpus
NVIDIA H100 GPU: https://www.nvidia.com/en-us/data-center/h100/

Overall, the choice of which system(s) to use for any process will depend on the specific use case. It doesn't make sense to buy the latest and greatest GPU from NVIDIA only to run the program that you built three years ago. Below we list our specific set-up for all our various use cases.

Our System: End-to-End

We currently utilize S&P Capital IQ data feeds for the bulk of our incoming data. We have fundamental financial data, stock price data, and financial filings delivered directly into our PostgreSQL database. We chose PostgreSQL due to its text, JSON capabilities, and the scaling flexibility it gave us over solutions like MS SQL Server, where we would have to pay for every additional core added to our database. We also use a few other publicly available APIs and datasets that are integrated at various levels in different applications. In addition to storing incoming financial data, we also store our internal data from the Data Science and Research teams in PostgreSQL. PostgreSQL extensions allow us to store a variety of data types such as text embeddings using 'pgvector' without having to connect multiple different databases together to store and access the same data.[39]

Our data pre-processing and predictive modules are a mixture of Python and Julia. We use Julia for intensive data processing and tabular machine learning. Python is mostly used for NLP and other deep-learning problems where we can utilize high-performance libraries such as 'PyTorch' that are wrapped in Python. By combining Python and Julia, we can efficiently develop pipelines to handle hundreds of gigabytes of data in Julia. This avoids the need for lower-level languages like Rust or C++. After processing, we then utilize Python's extensive ecosystem to train large models with this data.

While we have different outputs from different applications, the most important for us is the Talos 100 list. This list represents our system's current prediction of the 100 highest-quality companies in the world (for a given cap size) that Saguaro has never previously analyzed. As mentioned above, this list is given to the research team for idea generation to improve the efficiency and breadth of their search to find that next truly great business. Furthermore, we can leverage the list generation functionality to construct historical data sets. These sets and any other data we wish to consider can be put through

CPUs vs. GPUs

CPU – Central Processing Unit

- General Purpose
- Sequential Processing
- Fewer, but more powerful cores

GPU – Graphics Processing Unit

- Graphics Rendering
- Parallel Processing
- More, but simpler cores

GPUs outperform CPUs for machine learning tasks due to their superior memory bandwidth and parallel processing. These make GPUs significantly faster than CPUs while handling the huge number of matrix multiplications required for training neural networks.

For everything you would ever want to know about GPUs, please follow this link.

[39] pgvector: https://github.com/pgvector/pgvector

rigorous analysis using our dedicated backtesting library, 'SaguaroTrader.jl', to validate their importance or predictive power over time.[40]

Our hardware has grown since we purchased the first data science machine in April 2022. We started with a consumer-level desktop with an NVIDIA 3090.[41] In April 2023, we received our first server, a Scalar from Lambda Labs that contains four A6000 ADA GPUs (totaling 192 GB of GPU memory).[42] As our needs and capabilities scale, we will utilize the cloud to scale our GPU workloads.

We are happy to discuss our specific choices for data, software, or hardware in greater depth, and sadly note that the majority of what is written above will be obsolete within 48 months. Happy computing!

"Never bring a CPU to a GPU fight."

[40] https://github.com/SaguaroCapital/SaguaroTrader.jl - We'll communicate more about this library at a later date.
[41] NVIDIA 3090: https://www.nvidia.com/en-us/geforce/graphics-cards/30-series/rtx-3090-3090ti/
[42] Scalar from Lambda Labs: https://lambdalabs.com/products/scalar

How can we discern real AI from hype?

If all software is AI, then everyone can claim they're using AI, right? Yes and No. Anyone can purchase some off-the-shelf product that is powered by AI, but not everyone can develop bespoke tools that will add value. In this answer, we'll provide two distinct approaches to help discern real AI from hype. First, we outline some of the key principles that we believe lead to success with data science. Second, we provide a few questions that we would ask if we were sitting on the other side of the table.

Key Principles for Success

Cut-and-dried business problems: The philosophy of most AI projects is "throw it against the wall and see what sticks." Wandering aimlessly may lead to some discoveries, but it isn't going to move a business forward. The best technology or innovation teams we've interacted with start with the problem first and work backwards to find the technology they need. AI is nebulous enough that it is critical to begin with a clearly defined business problem that you're trying to solve.

Breaking it down: Given the current stage of AI, we are far from artificial general intelligence (AGI). Thus, large, complex problems must be broken down into manageable steps that existing models can handle. Modularizing AI tasks is crucial to achieving effective solutions. If anyone claims they have a single system that does everything end-to-end, they are either lying or don't understand their system.

Image created by Midjourney using the description *"the key principles for success."*

Putting your heads together: Data Scientists typically lack the domain expertise necessary to solve cut-and-dried business problems on their own. Therefore, open communication and tight collaboration between the data science team and the group with the problem to be solved is a must. In Saguaro's case, the participation of our research team is vital in driving the AI project forward. Open dialogue fosters a culture of shared learning and problem-solving, accelerating the project's timeline.

Pushing the envelope: There is no road to successful AI implementations not paved with numerous trials, errors, and learnings. A culture that fosters experimentation, and reasonable risk taking is a must. Both senior management and project leaders need to recognize that not every idea will pan out, and that that's okay. The key is to ensure that all mistakes still have positive net present value, i.e. the team learns from their mistakes.

Homegrown Data: Data forms the bedrock of AI. Machine learning models require data for any task. If you aren't gathering unique internal data, then it is hard to see how you are building something bespoke or truly differentiated. Our internally generated research data plays a crucial role in building comprehensive datasets, particularly for natural language processing (NLP) tasks.

Getting the bigwigs on board: Senior management's buy-in is paramount for the successful adoption and implementation of AI within an organization. Their endorsement drives resource allocation, strategic alignment, and overall momentum for any project. Not every company has AI in their DNA like Saguaro, but projects can succeed if those at the top are behind it.

Questions We Would Ask

How are you using AI in your process? We would ask this question of several individuals. If an organization truly uses AI, everyone should be able to describe its application similarly. Varied and imprecise answers indicate either a lack of understanding or a lack of real AI usage.

Can we see it run? Can we see the code? Real AI applications can be demonstrated. If an organization hesitates to show their system in action, it likely doesn't have one. The true value lies in the data and the code, neither of which can be stolen just by observing the output.

Can you give me an example of how AI has changed a process or how you do something? Obviously, we are hoping for an example, but more importantly, we are looking for enthusiasm. If an AI system adds value, everyone will be enthusiastic about it. Lack of excitement often indicates that the system isn't providing significant value.

How/Why? Can you explain that to me? Ask the next question to ensure that someone on the team has more than just surface level/cursory knowledge about the system. If someone builds something, they should be able to talk about it in detail/go back to first principles.

Why did you choose that language/package/hardware setup, etc? Be sure to ask something specific about some aspect of the underlying system to ensure that somebody knows how the backbone of the system actually functions.

How do you use randomization in your process/system? The answer is that you use it to reduce bias and overfitting during model training by intentionally introducing some randomness. You have to try some things that don't work to optimize the things that do. Or think of it like an automated pool cleaner that would be stuck forever in the corner without the occasional change of direction. We would ask some question like this to ensure they understand the basics.[43]

The primary eight questions in this primer should also be considered fair game. In conclusion, real AI is as much about the right approach and principles, i.e., embracing AI, as it is about the technology itself. Navigating the hype requires evaluating whether any firm is set up for success and then digging in with questions to discover if something of true value is there.

[43] Another example question: How would you solve a maze using an algorithm? For example, your algorithm could be to always follow the right wall of the maze. It could be to fill the maze with water and follow the path the water takes or any number of other systematic solutions. For a fun intro to this entire topic, check this out: Micromouse

How can a small firm compete with giants?

Often, it's assumed that small players cannot compete with their larger peers when it comes to technology. This assumption, we argue, is flawed. In our view, boutique managers have distinct advantages over passive and active industry giants, not just in investing, but also in the domain of data science and AI. We assert this for three reasons: 1) the industry's structure, 2) our previously explained philosophy of embracing AI, and 3) the innovator's dilemma. As previously discussed, AI is not a standalone advantage, but rather an amplifier. It merely enhances the unique value a firm already provides; it doesn't create value out of thin air. While the extent of this amplification depends on a firm's adeptness in utilizing AI tools, there is a limit. Regardless of a finance firm's resources, it is highly improbable that it will develop AI capabilities better than those offered by big tech. Let's discuss.

Industry Structure

Small boutique investment firms are not at a disadvantage when it comes to deploying AI, largely due to the structure of the AI industry itself. The giants of this technology such as Google, Microsoft, Amazon, and Open AI are the key developers of most AI tools and systems. They currently offer these tools for free, knowing that to use them well or at scale you'll have to adopt their broader ecosystems, which they monetize. This sets the stage for an investment firm battle not about who can build the most advanced tools, but rather who can best adapt and customize the pre-existing tools to their specific use case. Boutique firms, with their agility and niche focus, can tailor these tools to their particular needs in a way that massive, unwieldy industry behemoths may struggle to match. Do you believe BlackRock will build a custom system for each of their thousands of products?

Furthermore, it's important to note that the world's leading data scientists, programmers, and developers often contribute their best work to the open-source community. It's virtually impossible for any single company, even the biggest software firms, to outpace the innovation that stems from this vast network ceaselessly pushing the boundaries. The open-source community is an immense pool of talent and advancement; cutting yourself off from it to patent every last thing inevitably leads to subpar technology. It also makes recruiting top-tier talent that much harder. Small boutique firms, more nimble and able to adapt, can remain intimately connected with this community, absorbing, and implementing the latest advancements and recruiting those who want to give back. Here's an example of two open-source platforms from GitHub:

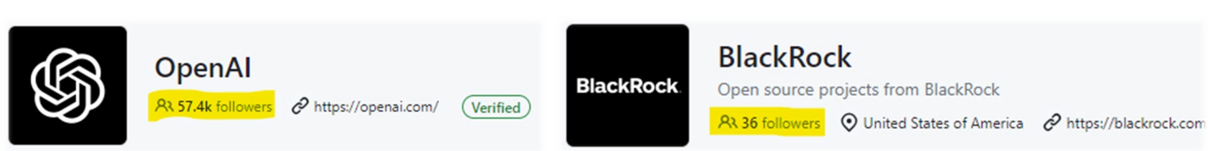


Image of OpenAI and Blackrock's Github account for illustrative purposes only
Photo illustration sourced from Github

Boutique firms have access to the same tools and the same open-source community as the largest industry giants. The difference is who utilizes them better.

Embracing AI: Speed of Adoption, Idiosyncratic Data, and Firm Adaptability

Considering AI's constantly evolving landscape, we've advocated for a philosophy of embracing AI as described above. This embrace is a holistic approach to understanding and employing AI technologies. This approach accepts that no single AI technology or set of technologies can maintain permanent relevance. It anticipates the inherent flux within the AI sphere and encourages continual adaptiveness to its ceaseless progression. To illustrate, current technology such as ChatGPT 4.0 might be at the forefront today, but within a year the tech community will have progressed to the next generation. Still, the workflows or processes that utilized such technology will endure. AI in this respect is an amplifier rather than a standalone differentiator for businesses. Therefore, the success of an investment firm hinges not solely on employing AI - which is inevitable - but rather on the unique aspects of that firm and its philosophical approach towards incorporating AI into its processes. Key to this philosophy are three elements: 1) speed of adoption, 2) quality and quantity of unique data, and 3) the capacity to reshape the firm's processes to optimally leverage these technologies.

In terms of the first element of embracing AI—speed of adoption—small boutique investment firms hold a distinctive edge. Their nimble structure often allows them to adapt quickly, integrating new technologies before they become mainstream, and thus potentially outpacing larger industry players who may struggle with bureaucratic inertia. Just as early adopters of sentiment analysis reaped significant profits years ago, the small firms with the agility to quickly implement nascent technologies can achieve a unique competitive edge. For years we've created competitive edges by building tools years before they were commercially available. In this realm there are no permanent moats.

The second element—quality and quantity of unique data—is another area where small firms can excel.[44] Large firms may have access to vast data sets, but this data is often homogenous, widely available, and not focused on specific strategies. Boutique firms, on the other hand, can focus on cultivating unique and idiosyncratic data sources that feed into their specialized strategies, providing them with insights that large firms, operating on a broader scale, might overlook. High-paid research analysts and portfolio managers are often reluctant to reformat their work to make it machine-readable, a process that greatly enhances data collection efforts. The edge here lies not in the sheer volume of data, but in its uniqueness and relevance to the firm's particular investment strategy.

Lastly, regarding the ability to reshape a firm's processes to optimally leverage AI technologies, smaller firms have the advantage of flexibility. Larger organizations often face more significant challenges when trying to implement substantial process changes due to their size, complexity, and pre-existing investment mandates. In contrast, boutique firms, with less red tape and more streamlined decision-making processes, can more easily redesign their workflows, integrate AI technologies more seamlessly, and thus optimize their use. This inherent flexibility allows small firms to continually adapt their processes in response to AI's relentless progress, ensuring they stay competitive in a rapidly evolving technological landscape. The philosophy of embracing AI favors boutique investment firms over industry giants in a way that is self-evident.

---

[44] Please note that we are limiting this discussion to long-term investing. We are not suggesting that boutique firms can beat Citadel or Renaissance Technologies at their own game. Ren Tech's monthly electric bill alone exceeds our annual budget.

Innovator's Dilemma

In addition to the above philosophy favoring smaller firms, the industry giants have to contend with the innovator's dilemma. Those with substantial passive investment operations, which meet existing market demands and that provide consistent revenue, might be skeptical of AI-driven products. While new AI products might meet a future market need, they also carry the risk of failure and financial loss. These AI-driven investments are also likely to be more expensive to operate, which could deter firms that have focused on nothing but lowering costs for the last twenty years. Small firms, and especially start-up firms, do not face this headwind.

While large firms certainly have advantages, namely resources and brand, their size and complexity can also be a liability. Given the industry structure, their ability to embrace AI, and the innovator's dilemma facing the incumbents, we believe that smaller firms are well-positioned to make significant inroads in the application of AI to investment management.[45]

"It's not the size of the data that matters, but how you use it."

[45] The fact that we've staked our careers and reputation on this opinion, by starting our own boutique investment firm, should tell you all you need to know about both our conviction and bias.

How will AI change investing by 2030?

After writing many words, please allow us one bold prediction: the coming decade holds more change than the previous century. Now that AI has entered the investing world, our pace of change becomes exponential. All firms must become technology firms, even fundamental equity shops. Here's how we envision this future:[46]

1. A Shift to Soft and Idiosyncratic Data

In less than five years, we predict that AI will consistently outperform humans in analyzing and drawing insights from widely available text and financial data. As a result, human research roles will pivot towards reading "between the lines," searching for idiosyncratic data like an investigative journalist, seeking out new data sources, and checking machine-output for errors.

2. Human-Machine Collaboration

AI tools like Chat GPT will continue to permeate professional settings. Companies like Microsoft are already integrating AI-driven functionalities into products like Excel and Word, suggesting that it won't be long before research teams collaborate directly with AI tools in their processes. Human and machine outputs will no longer be distinct but will become increasingly intertwined, providing deeper, more nuanced insights.

3. Internal Data Capturing and Labeling

To fully utilize AI, firms will lean into capturing their own internally generated data: every meeting, every call, every discussion. Not only will this lead to better training data for AI tools, but it will also foster a more informed and effective communication flow. There is a maxim: "what gets measured, gets done." We believe that measurement is a pre-requisite for continuous improvement over the long run. Those seeking to improve will have ample feedback, and those who aren't may have a more uncomfortable performance review. This effort to systematically capture, label, and understand internally generated data will become a defining characteristic of successful investment firms.

4. Seamless Integration of AI into Routine Tasks

AI will revolutionize how we interact with standard software tools. Imagine interacting with your spreadsheet, having a conversation about the data you need, the model you're constructing, or the points you want to highlight from an earnings call. This future is closer than we might think. Even today, most standardized programming tasks are being performed by AI, not humans. And we all know at least one person in the office who will have one of these early next year.[47]

5. Exploitation of Long Tail Information

AI's ability to process vast quantities of data will allow it to explore the 'long tail' of information. This includes sources such as 8-Ks, news articles, and social media feeds. Reviewing all of these will offer

[46] For a more general prediction of AI impacting humanity, please read the latest from the great Marc Andreessen: https://a16z.com/2023/06/06/ai-will-save-the-world/

[47] For those reading, we linked Apple's new vision pro: https://www.apple.com/apple-vision-pro/

new insights into companies and industries. This comprehensive parsing of information could lead to earlier identification of disruptions or market shifts and will likely become table stakes for any professional investor. The great Robert Caro once defined his research philosophy with a simple phrase: “Turn every page.”[48] Investment firms will be able to finally live out this mantra on every company.

6. Transition from Data Overload to Insight Generation

While data is increasing exponentially, what investors need is not more data but more insights. AI will assist by processing large volumes of information and highlighting key insights, making the work of analysts and portfolio managers more focused and efficient. Imagine a global macro-economic analysis performed daily, but summed up in a single sentence or paragraph that could direct attention where needed.

7. New Focus on Data-Rich and Data-Poor Companies

The paradigm of "over-followed" or "under-followed" companies will be replaced by "data-rich" or "data-poor" companies. Data-rich companies will have their operations transparently analyzed by AI, whereas data-poor companies will become the inefficient markets of the 21st century, offering a ripe ground for analysts who can uncover valuable hidden insights.

8. AI Regulation Management

AI will automatically interpret, monitor, and ensure compliance with complex and ever-changing financial regulations, reducing the risk of human error and mitigating legal risks. We predict that allocators will eventually demand AI-compliance agents working on any product in which they invest. We hope this happens soon.

9. Analysis of Corporate Transparency and Truthfulness

The assessment of corporate transparency and management honesty will be drastically enhanced. AI will be able to suggest when humans are evading or being less than truthful. It will also scrutinize company communications including earnings calls and public statements for consistency and truthfulness. By comparing management promises against actual results and analyzing executive language for signs of evasion or ambiguity, AI could provide valuable insights. This could foster more informed investment decisions and promote a more transparent business environment.

10. The Rise of “Passintelligence” Funds

Combining the best elements of both active and passive management, these AI-powered funds will adapt in real-time to market changes and make intelligent investment decisions. They will offer broad diversification, minimal fees, and unlike their passive brethren, they will contribute to markets functioning properly. These funds will also be infinitely customizable, need we say more? With

[48] Robert Caro is a Pulitzer Prize-winning American journalist and author, best known for his extensive biographies including "The Power Broker" about urban planner Robert Moses, and his ongoing series "The Years of Lyndon Johnson." Born on October 30, 1935, Caro's meticulous research and powerful narratives have established him as one of the most respected biographers of his generation, known for shedding light on the exercise of political power.

Passintelligence funds, the best of active and passive will meet unparalleled customization, defining the future of investment management.

And many more...

AI is poised to dramatically reshape investment management over the next decade. Although we can't fully predict what this will look like, it's clear that firms that start their AI journey today will be better positioned in five years than those that hesitate. The transition to AI is not just about adopting a singular new technology; it's about a fundamental shift in approach and practice.[49] For firms that are already unique and add value, AI will be a powerful amplifier. For others, it may not be the silver bullet they hope for. One thing, however, is certain: doing nothing is not an option in the new AI-driven era.

"Generative AI is the most powerful intellectual tool in history, and it is improving exponentially."

[49] Please see the embracing AI philosophy described above.

What are your top AI resource suggestions?

We've broken these resources into six categories: overview, history, philosophy, examples, tools, and future. Much more than just sources for the paper above, though we've cited many in the primer itself, we hope these suggestions can be steppingstones on your journey of discovery for AI, computer science, and their potential application to investment decision-making.

Overview – Some resources to help get you up to speed in general:

1. Harris, D., Bornstein, M., & Appenzeller, G. (2023, May 25). AI Canon. Andreessen Horowitz.
   Article Link
   You must visit this page. It contains even more resources than we've provided here. The best resource we know of to get up to date. AI Canon is a comprehensive collection of essential knowledge and trends about the field of artificial intelligence, curated by Andreesen Horowitz.

2. Logston, D., & Ferreira Filho, W. (2017). *Computer Science Distilled: Learn the Art of Solving Computational Problems*. Code Energy LLC. Amazon Link
   If you want (or need) to start at the very beginning, this is the best book we've found. It is short, easy-to-read, and potentially life changing. The book is a practical guide for beginners and enthusiasts seeking to develop understanding and skills in the large field of computer science.

3. Harvard University. (2023). CS50's Introduction to Computer Science. edX. Course Link
   If you prefer learning in a classroom, lecture-type format. This is our preferred course on computer science. It isn't easy, but you'll be 80% of the way there at the end (and ahead of most of us).

4. Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
   Amazon Link
   This is where to start for a true base education in AI. A widely acclaimed textbook written by the famed Stuart Russell and Peter Norvig, it provides a comprehensive overview of the field of artificial intelligence. This is the AI Bible.

5. Ng, A. (2023). Machine Learning Collection. Coursera. Course Link
   These are the famous AI and machine learning courses taught by the renowned Andrew Ng. Still the gold standard for learning the basics of modern AI and machine learning.

6. Benaich, N., & Hogarth, I. (2022). State of AI Report 2022. State of AI. https://www.stateof.ai/
   We repost this report every year as we find it the most comprehensive update on the current state of AI. Value add for novice and professional alike.

History – If you are a history buff like we are, here are few good ones:

1. Press, G. (2016, December 30). A very short history of artificial intelligence (AI). Forbes.
   Article Link
   If our ultra-brief history of AI isn't enough, Gil Press offers a slightly longer history of AI and how we got to where we are today.

2. Wooldridge, M. (2021). A brief history of artificial intelligence: What it is, where we are, and where we are going. Flatiron Books. Amazon Link
For even more, this offers a comprehensive exploration of the evolution of AI, its current state, and potential future trajectories without the need of a technical education.

3. McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133. Paper Link
This is the original paper in computational neuroscience. It proposed a mathematical model of how neurons in the brain function and paved the way for modern neural networks and modern AI systems.

4. Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460. Paper Link
This ultra-famous paper from Alan Turing explores the concept of machine intelligence and poses the question, "Can machines think?" The core of the paper introduces the Turing Test, which evaluates a machine's ability to exhibit human-like intelligence.

5. Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J., & Ng, A. Y. (2012). Building high-level features using large scale unsupervised learning. In Proceedings of the 29th International Conference on Machine Learning (ICML-12) (pp. 81-88). Paper Link
This is the paper that set off the modern unsupervised learning craze. It is renowned for showcasing that a machine learning model could self-learn to detect complex features like faces without needing labeled data. It is also famous due to the prominence of its authors in artificial intelligence research.

6. SRI International. (2017, February 23). Shakey the Robot: The First Robot to Embody Artificial Intelligence. YouTube. YouTube Link
Fun video on Shakey the Robot, the first AI robot. An influential and groundbreaking project in the field of artificial intelligence developed at the Stanford Research Institute (now known as SRI International), Shakey was crucial in shaping the field of AI and serving as a foundation for subsequent developments in robotic systems, perception, planning and reasoning.

Philosophy – These books shaped our philosophical approach towards technology, AI, and more:

1. Dalio, R. (2017). *Principles: Life and Work*. Simon & Schuster. Amazon Link
Principles of Life and Work showcases how Dalio developed his own mental models, including the concept of the "economic machine" and transformed them into a systematic, algorithmic process. It emphasizes the importance of understanding and adapting your own processes to navigate and succeed in the business world.

2. Hofstadter, D. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books. Amazon Link
This book is not for everybody, but for some it is an epiphany. Hofstadter uses art, logic, and music as metaphors to express a philosophical approach to AI, suggesting that true intelligence and self-consciousness might emerge from well-structured systems of

meaningless elements, just as the complex structures and patterns in Bach's music, Escher's drawings, and Gödel's mathematics emerge from simple elements. It advocates for the view that AI, much like human intelligence, isn't just about mimicking or understanding the formal rules, but rather about discovering and revealing self-reference, meaning, and "strange loops" within the system. You try summarizing this Tour de Force with fewer words!

3. Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux. Amazon Link
   This book likely doesn't need an introduction but must be included on the list. Thinking Fast and Slow draws on decades of research in psychology and behavioral economics to demonstrate how the two systems, intuitive and automatic (system 1) and the deliberate and analytical (system 2), shape our judgement and choices. The book reveals cognitive biases and heuristics that often lead us astray and discusses how to build processes, including AI, around them.

4. Montier, J. (2010). *The Little Book of Behavioral Investing: How Not to Be Your Own Worst Enemy*. John Wiley & Sons. Amazon Link
   One of our favorites, The Little Book of Behavioral Investing: How Not to Be Your Own Worst Enemy, delves into the world of behavioral finance. Montier presents a compelling argument for the significant impact of human behavior on investment decisions, discussing psychological biases, cognitive errors, and how to avoid them.

5. Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press. Amazon Link
   The Innovator's Dilemma explores how successful companies can become vulnerable to disruption. The book examines why established companies often fail to adapt to disruptive technologies due to their focus on existing customers and business models. It emphasizes the importance of recognizing and addressing disruptive innovation through separate organizational structures and strategies to maintain long-term success in dynamic markets.

6. Weaver, R. M. (1984). *Ideas Have Consequences*. University of Chicago Press. (Originally published in 1948) Amazon Link
   The granddaddy of them all. Ideas Have Consequences explores the relationship between ideas and their impact on society, culture, and individuals. Weaver argues that ideas shape the way we think, perceive the world, and act. This book challenges readers to critically examine the ideas they embrace and to consider the long-term implications of those ideas on society and their individual lives.

Examples – If you'd like further examples of AI in Investment Management, turn here:

1. Bartram, S. M., Branke, J., & Motahari, M. (2020). Artificial intelligence in asset management. CFA Institute Research Foundation. Research Link
   The CFA Institute Research Foundation discussed the applications of AI in various parts of asset management. They discuss the trends that they see occurring in the industry, and common applications of AI within the various parts of the industry.

2. Cao, L. (2023). Handbook of artificial intelligence and big data applications in investments. CFA Research Foundation Books. Research Link
   The CFA Institute's latest compendium on AI in the Investment Industry, this book is nothing but examples of AI being applied in every niche of the investment world.

3. McGranahan, C., Parker, M. W., & Huang, J. (Eds.). (2023). Artificial Intelligence: Getting Smarter. Bernstein.
   We apologize for the lack of link, but this is Bernstein's latest thinking on how generative AI is likely to impact every industry in the world. A must read for investment practitioners.

4. Wu, Shijie, et al. "BloombergGPT: A large language model for finance." arXiv preprint arXiv:2303.17564 (2023). Paper Link
   BloombergGPT is a 50-billion parameter casual language model. It was pre-trained using Public text datasets combined with Bloomberg's internal financial text dataset, FinPile, which consists of 363 billion tokens of financial text. This is Chat GPT for finance, coming soon.

5. Capponi, A., & Lehalle, C. -A. (Eds.). (2023). *Machine learning and data sciences for financial markets: A guide to contemporary practices*. Cambridge University Press. Book Link
   This is the latest compendium of machine learning applications in the world of finance. We are currently reviewing the book ourselves for ideas. Please be warned that this is advanced material which assumes significant knowledge of both math and various ML techniques.

Tools – Our tool suggestions can be found above, but here are few other helps:

1. Dettmers, T. (2023, January 30). Which GPU(s) to Get for Deep Learning: My Experience and Advice for Using GPUs in Deep Learning. Tim Dettmers. Article Link
   GPUs are a fundamental part of AI and have enabled recent advances to be computationally feasible. NVIDIA and its CUDA ecosystem are the drivers behind the massive adoption of GPU machine learning. This blog by Tim Dettmers goes into depth about how GPUs work and why GPUs are great for training neural networks.

2. Rush, S., & Weiss, G. (n.d.). Thinking Like Transformers. Sasha Rush. Retrieved (2023, February 3), from Paper Link
   Thinking Like Transformers gives a visual introduction to Transformer models. It is a combination of Python code to implement the barebones of a Transformer and diagrams showing what the code is actually doing.

3. Lambert, N., Castricato, L., Von Werra, L., & Havrilla, A. (2022, December 9). Illustrating reinforcement learning from human feedback (RLHF). Hugging Face – The AI community building the future. Article Link
   This blog post introduces the concept of RLHF and provides a jumping off point for readers that want to dive into literature on the topic.

4. Lundberg, S., Lee S. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems 30,* 1-9. Paper Link

Scott M. Lundberg and Sun-In Lee at the Paul G. Allen School of Computer Science introduce a unified framework for interpreting model predictions, SHAP. SHAP assigns feature importance for a single observation to help interpret what features had the largest impact on the prediction.

5. Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. University of Florida - Department of Finance, Insurance and Real Estate. Paper Link
This ultra-popular paper outlines that ChatGPT may do a better job predicting the sentiment of headlines and articles than any previous approach and that this method generated excess returns at the time of publication. Cutting-edge.

6. Yang, H., Liu, X.-Y., & Wang, C. D. (2023). FinGPT: Open-Source Financial Large Language Models. Paper Link
This is the world's first open-sourced finance-focused Large Language Model (LLM). You can believe that we are spending time here. Cutting-edge.

Future – If you, like us, are most excited about the future possibilities of AI, we have you covered:

1. Andreessen, M. (2023, June 6). Why AI will save the world. Andreessen Horowitz. Article Link
In his persuasive article, 'Why AI Will Save the World', Marc Andreessen addresses the prevalent concerns about the potential negative impact of Artificial Intelligence on society. Through his reasoned analysis, Andreessen argues that apprehensions surrounding AI often tend to be overstated. He posits that the pursuit of AI technology is unlikely to yield catastrophic consequences, presenting instead an optimistic vision of AI as a force for good.

2. Daugherty, P. R., & Wilson, H. J. (2018). Human + Machine: Reimagining Work in the Age of AI. Hardcover edition. Amazon Link
Human + Machine argues that the AI paradigm shift is a transformation of all business processes within an organization. Humans and machines will continue to work more collaboratively over time, which will allow business processes to become more fluid and adaptive.

3. Kurzweil, R. (2000). The Age of Spiritual Machines: When Computers Exceed Human Intelligence. Penguin Books. Amazon Link
Ray Kurzweil predicts that AI will surpass human intelligence and become increasingly integrated into human activity. He sees a future where machines may gain consciousness, have spiritual experiences, and transform our world.

4. Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking. Amazon Link
The Singularity is Near: When Humans Transcend Biology takes readers through a journey into the future where artificial intelligence, genetics, nanotechnology, and robotics reshape every facet of our lives. Kurzweil presents a thought-provoking vision drawing on cutting edge

research and his deep expertise in technology to ask the question, can we transcend our biological limitations?

5. Urban, T. (2015, January 22). The AI Revolution: The Road to Superintelligence. Wait But Why.
   Article Link
   The AI Revolution: The Road to Superintelligence explores the potential impact of artificial intelligence on society and the concept of superintelligence. Tim Urban delves into the capabilities and potential risks associated with AI on three levels: narrow AI, general AI, and superintelligence.

6. Urban, T. (2015, January 27). The AI Revolution: Our Immortality or Extinction. Wait But Why.
   Article Link
   The AI Revolution: Our Immortality or Extinction is a follow up article in the AI Revolution series. Here, Urban covers the possibility of AI superintelligence in our lifetime and what it means to have speed superintelligence versus quality superintelligence. The article provides a peek into what AI could mean for the future of humanity.

"Copy from one, it's plagiarism; copy from two, it's research."

An Ultra-Brief History of Artificial Intelligence

1943 – Warren McCulloch and Walter Pitts pen the original mathematical model for how neurons may function in the human brain, paving the way for today's neural networks.

1950 – Alan Turing proposes the Turing Test, a measure of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.

1952 – Claude Shannon built Theseus, a maze-solving mouse that used a bank of relays to solve the problem. A mechanical precursor of later systems, Shannon's work revolutionized the telephone system and led to the development of the modern computer.[50]

1956 – John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon coin the term "Artificial Intelligence" (AI). The Dartmouth Conference takes place, marking the birth of AI as a field of study.

1957-1974 – The "Golden Years" saw heavy government investment in AI and were characterized by great optimism and significant achievements. This era laid the foundation for many techniques and approaches used in modern AI.

1966 – Shakey, the first AI-based robot, was introduced. It could move around its environment, perceive its surroundings, and plan its actions.

1974-1980 – The first "AI Winter." Funding was reduced as AI failed to live up to early expectations.

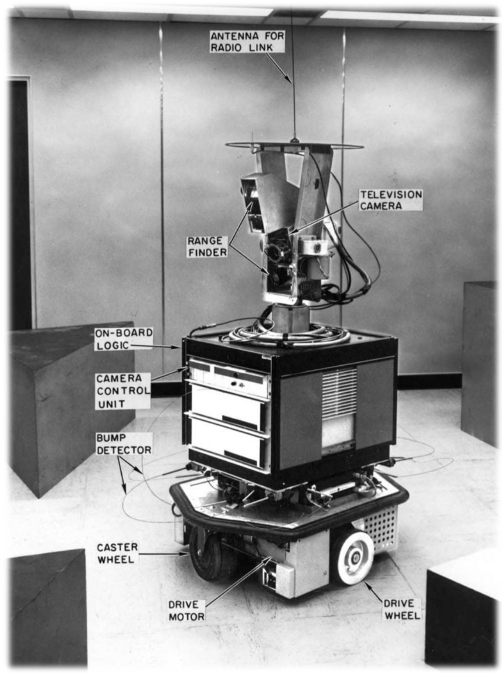


Image of Shakey the robot the first AI based robot.
Image sourced from The New Stack

1980-1987 – Development of expert systems, which use if-then logic to solve problems leads to a brief revival in the field.

1987-1993 – The second "AI Winter." The limits of expert systems were exposed.

1989 – Yann LeCun developed the first practical application for backpropagation, the mechanism by which modern neural networks learn.

1997 – IBM's Deep Blue beats World Chess Champion Garry Kasparov.

2005 – Stanley, a robotic vehicle developed at Stanford, wins the DARPA Grand Challenge, a long-distance race for driverless cars.

---

[50] https://webmuseum.mit.edu/detail.php?module=objects&type=related&kv=76066

2011 – IBM's Watson wins Jeopardy against two of the show's record winners.

2012 – Google's X Lab develops a machine learning algorithm that can identify cat videos with a high degree of accuracy.

2012 – AlexNet kicked off the deep learning GPU revolution by topping countless image recognition benchmarks with a model trained for five to six days on two NVIDIA GTX 580 3GB GPUs.

2014 – The Attention Mechanism behind Transformers was introduced by Dzmitry Bahdanau.[51]

2015 – Google's DeepMind develops AlphaGo, which defeated World Go Champion Lee Sedol. Based on unsupervised learning, this result was achieved years ahead of schedule.

2017 – Google releases the "Attention is All you Need" paper starting the current Transformer model craze.[52]

Image of Google's AlphaGo versus world champion Lee Sedol
Image sourced from The Guardian

2020 – OpenAI releases GPT-3, which could create human-level text given a simple prompt.

2020 – Google releases AlphaFold2, which predicts protein structures with unprecedented accuracy.

2021 – OpenAI unveiled DALL-E, a revolutionary model that generates images from textual prompts. This groundbreaking development was later complemented by OpenAI's DALL-E 2 and Runway's open-source counterpart, Stable Diffusion.

2022 – OpenAI releases Chat GPT, a chatbot interface to GPT-3.5 that has been aligned using reinforcement learning from human feedback (RLHF).

2023 – OpenAI's GPT-4, Google's PaLM 2, and Facebook's LLaMA are spearheading advancements in the field of large language models. The Generative AI era has begun.

"Algorithms that forget their history are doomed to repeat it."

[51] https://arxiv.org/abs/1409.0473
[52] https://arxiv.org/abs/1706.03762

Glossary (Alphabetical)

Activation Function: The math used in each neuron of a neural network that determines if that neuron fires or "activates" based on the inputs received. Neural networks can learn complex patterns because this math is unique to each neuron. More technically this math introduces non-linearity to the model, it works by calculating the weighted sum and adding bias. Common examples include Sigmoid, ReLU, and Tanh.

Algorithm: A step-by-step procedure or set of rules designed to perform a specific task or solve a particular problem, often used in computing and data processing.

API: An Application Programming Interface (API) is a set of rules and protocols that allows different software applications to communicate and interact with each other. It defines how software components should interact and enables developers to access the functionality of an existing application or service in a standardized way, without needing to understand the underlying implementation details.

Array: An array, as a basic data structure in computer science, is a collection of elements or values, each identified by one or more indices (the position of the element in the array). Think of an array as a row of lockers. Each locker is of the same size and shape, and they can each hold one item. Just like you'd need to know the number of a locker to find what's inside, you need to know the position of an item in an array to access it. These elements are usually of the same data type—such as integers, floats, or strings—and are stored at contiguous memory locations. This means that the elements are placed next to each other in memory, which allows for efficient access and manipulation of the data. The dimensionality of an array refers to how many indices are used to access an individual element. In other words, it indicates the number of directions in which the array extends.

Artificial General Intelligence (AGI): A type of artificial intelligence that possesses the capability to understand, learn, adapt, and implement knowledge across a broad range of tasks at a level equal to or beyond human cognitive abilities. Unlike narrow AI, AGI can transfer learning from one domain to others autonomously, exhibiting cognitive flexibility and independent problem-solving skills.

Artificial Intelligence (AI): The field of computer science that aims to create machines capable of performing tasks that would normally require human intelligence, including, but not limited to problem-solving, learning, understanding natural language, speech recognition, and visual perception.

Attention Mechanism: A computational model used in deep learning architectures to assign varying degrees of importance to different inputs or features. By allowing the model to focus on relevant parts of the input for a given task, the attention mechanism enhances the model's ability to handle complex tasks, manage long sequences, and capture dependencies in the data. This technique has been crucial in achieving state-of-the-art results in tasks such as natural language processing, image recognition, and more.

Autonomous System: A type of AI system that can perform tasks or make decisions independently, without human intervention by integrating perception, reasoning, learning, and decision-making capabilities.

Backpropagation: An algorithm in machine learning, specifically in neural networks, used to calculate the gradient of the loss function with respect to the model's weights. It optimizes the weights by propagating the error backwards from the output layer to the input layer.

Bias: The systematic error introduced by an algorithm that leads to incorrect predictions. It's often due to oversimplification of a model or assumptions in the learning algorithm.

Bigram: In the context of Natural Language Processing (NLP), a bigram is a sequence of two adjacent elements in a string of tokens, typically words, taken from a larger corpus. The elements are not necessarily characters, they could also be words, sentences, or other units depending on the application. Bigrams are a subset of n-grams, where n represents the number of units in a sequence. In a bigram, n is 2. For example, in the sentence "I love to play soccer," the bigrams would be: ["I love," "love to," "to play," "play soccer."]

Casual Language Model: A language model that generates text by predicting the next word or token in a sentence based on the words or tokens that came before it. These models are "causal" because the predictions they make are dependent on the preceding sequence of words, which mirrors the way humans typically generate sentences in a "cause and effect" manner.

Code Wrapper: A way to interact with a piece of code written in a different language from the one you are using. The wrapper essentially "translates" between the two languages, enabling them to communicate with one another.

Computer Vision (CV): A subfield of AI that focuses on the development of techniques and algorithms that enable computers to interpret and understand visual information from the surrounding world, such as images or videos.

Deep Learning (DL): A subset of ML that involves the use of artificial neural networks, particularly deep neural networks, for modeling and analyzing data with a hierarchical structure. This allows machines to learn representations of data with multiple levels of abstraction.

Dimensionality Reduction: A technique used in AI and ML to reduce the number of input variables (features) in a dataset, often by transforming the data into a lower-dimensional representation that retains most of the relevant information. The goal is simplifying the data, reducing computational complexity, and mitigating the curse of dimensionality. Dimensions = Axes. Fewer dimensions = fewer coordinates or inputs.

Domain Expertise: The comprehensive knowledge and proficiency in a specific, well-defined area or subject, often acquired through extensive study, research, or professional experience. This expertise enables individuals or systems to solve complex problems and make informed decisions within that particular domain.

Expert Systems: A type of AI program designed to solve complex problems by reasoning through bodies of knowledge using if-then rules rather than conventional procedural code. The hope was for these to mimic the decision-making abilities of a human expert in a particular field.

Feature Engineering: The process of selecting, transforming, or creating relevant input variables (also known as features) from raw data to improve the performance of ML algorithms, often by incorporating domain knowledge or exploiting relationships among data attributes.

Feature Selection: The process of identifying and selecting the most important and relevant input variables (features) from a dataset to be used in the construction of an ML model, with the goal of reducing the complexity of the model, improving its interpretability, and minimizing overfitting.

Function: In machine learning, a function often serves as a mathematical model for predicting outputs based on inputs. It maps the input features to the predicted output. These functions can be as simple as a linear equation or as complex as a deep neural network. For instance, consider a simple linear regression problem where we want to predict a person's weight (y) based on their height (x). We could use a function like '$y = mx + c$', where 'm' is the slope and 'c' is the y-intercept. This function represents a model that assumes a linear relationship between height and weight. In more complex models like neural networks, the function becomes more intricate. The input is passed through multiple layers, each of which applies its own function to the input it receives and passes the result to the next layer. This can be seen as a composition of many functions.

Gradient: In the context of machine learning and optimization, the gradient of a function gives us the direction in which the function increases most quickly. For a loss function in machine learning, we typically want to minimize the loss, so we often move in the direction opposite to the gradient, known as gradient descent. In simple terms, if you imagine standing on a hillside and the hill represents your function, the gradient would tell you in which direction to go to ascend the hill as quickly as possible. To descend, you'd want to go in the opposite direction.

Gradient Descent: See Gradient.

Graphics Processing Unit (GPU): A specialized electronic circuit designed to handle graphics and images. It can quickly perform parallel simple math calculations to produce images. Machine learning, and especially deep learning, typically involve massive amounts of data and high-dimensional mathematical operations that can be parallelized, which makes GPUs a perfect fit for these tasks.

Hidden Layer(s): These are the layers between the input and output layers where the network learns to extract useful features and patterns from the input data. Each node in a hidden layer represents a learned feature, and the connections between layers represent how these features are combined to create new, higher-level features.

Input Layer: The initial layer in a neural network where raw data is fed into the model. It directly receives input features and passes them to the subsequent hidden layers for further processing. The number of nodes in the input layer typically matches the number of features in the data.

Large Language Models (LLM): Machine learning models trained on a vast amount of text data. They can generate human-like text by predicting the next word in a given sequence of words. The models 'learn' patterns and structures of the language through their training data and can answer questions, write essays, summarize texts, translate languages, and even generate creative content like poems or stories. Examples include GPT-3 and BERT models.

Layer: In the context of neural networks and deep learning, a layer is a collection of nodes or neurons through which data passes in a structured way. Each layer performs specific operations on its inputs and passes its outputs to the next layer. These operations typically involve weighted sums and activation functions. There are several types of layers, including input layers, hidden layers, and output layers, each serving a distinct role in the network.

Loss Function: Also known as a cost function or error function, this is a method used in machine learning to measure how well a machine learning model is performing or how far off its predictions are from the actual results. It quantifies the discrepancy between the predicted and actual values in a numerical way. In a nutshell, the goal of training a machine learning model is to find parameters (weights in the case of neural networks) that minimize the loss function. Different types of loss functions are used depending on the specific type of problem being solved.

Machine Learning (ML): A subfield of AI that focuses on the development of algorithms and statistical models that enable computers to improve their performance on specific tasks through the identification of patterns within data, without the need for explicit programming.

Metadata: The information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage data. It's essentially data about data, such as the author, date created, and format of a document.

N-gram: See Bigram.

Narrow Artificial Intelligence (NAI): A type of artificial intelligence designed to perform a specific task or a set of closely related tasks, such as voice recognition or image analysis. While often highly proficient within its designated domain, narrow AI lacks the capability to generalize its knowledge or abilities to new tasks or broader contexts, distinguishing it from Artificial General Intelligence (AGI).

Natural Language Processing (NLP): A subfield of AI that focuses on the interaction between computers and human languages, allowing machines to understand, interpret, and generate text or speech in a manner that is both meaningful and contextually relevant.

Neural Network (NN): A computational model inspired by the structure and function of biological neural networks, consisting of interconnected nodes (also known as neurons) organized in layers, which process information through a combination of weighted inputs, activation functions, and bias terms.

Node: In the context of neural networks, a node, also commonly referred to as a neuron or unit, is a fundamental building block of the network. Each node in a neural network takes in some input, applies some function to this input, and then passes the output to the next layer of nodes (unless it's an output node).

Overfitting: A phenomenon in which an ML model learns to perform exceptionally well on the training data but performs poorly on new, unseen data, often due to the model capturing noise or random fluctuations in the training data rather than the underlying patterns or relationships.

Output Layer: The final layer in a neural network that produces the result for given inputs. It translates the processed features from the hidden layers into predictions that align with the desired format, such as class labels for classification or numeric values for regression.

Pretraining: The process of training a neural network on large amounts of training data before fine-tuning the model for a specific task. This pretraining can come in many forms, and some common NLP examples are masking out random words then having the model predict the words, and predicting the next token given an input text.

Probability Distribution: A mathematical function that describes the likelihood of different outcomes in a given event or experiment. Probability distributions always sum to one.

Programming Language: A formal language comprising a set of instructions that produce various kinds of output, used in computer programming to develop software, applications, scripts, or other sets of instructions for computers to execute.

Randomization: The process of reducing bias and overfitting in model training by intentionally introducing randomness. This can involve randomly shuffling data before splitting it into training and testing sets, initializing weights randomly in a neural network, or employing stochastic algorithms. The aim of randomization is to ensure generalization and prevent models from learning spurious correlations.

Reinforcement Learning (RL): A method of ML where an agent learns to make decisions by interacting with an environment, receiving feedback in the form of rewards or penalties, and adjusting its actions accordingly to maximize cumulative reward over time.

Reinforcement Learning from Human Feedback (RLHF): A method of ML combining RL with human feedback. It involves an agent learning from trial-and-error interactions with a dynamic environment, augmented by feedback from a human supervisor. The human feedback helps to guide the learning process, correcting mistakes and influencing the agent to make optimal decisions.

Relational Database: A relational database is a type of database that organizes data into tables, which can be linked based on data common to each. Each row (also known as a record) represents a single entry in the database, and each column represents a type of data that may or may not be linked to other columns in the database.

Sentiment Analysis: An NLP technique that aims to identify, extract, and quantify subjective information, such as opinions, emotions, or attitudes, expressed in textual data, often used for monitoring public sentiment or customer feedback.

Singularity, The: A hypothetical future point in time when artificial intelligence will have progressed to the point of a greater-than-human intelligence, radically changing civilization, and potentially leading to rapid technological growth beyond human comprehension or control.

SQL: Structured Query Language (SQL) is a standardized programming language used for managing and manipulating relational databases.

Structured Data: Any data that resides in a fixed field within a record or file. Spreadsheet data and tables stored in relational databases are common forms of structured data.

Supervised Learning: A method of ML in which a model is trained using a dataset containing input-output pairs. The machine then creates a model which would transform the inputs into the provided outputs. The goal is that this model can be used to make predictions on new, previously unexamined data.

Tabular Data: Data that is organized in a table format with rows and columns, often used in spreadsheet or database applications. Each row represents an individual record or data point, and each column represents a specific field or attribute of the data. Examples include Excel spreadsheets or SQL database tables.

Text Classification: A process in Natural Language Processing (NLP) wherein text data is automatically categorized into predefined groups, often with the aid of machine learning algorithms. Applications include spam detection, sentiment analysis, and topic labeling.

Technepareia: From the Greek roots “techne” (art, skill) and “pareia” (readiness, ease), it signifies a readiness and commitment to learn about and utilize the continually evolving collection of technologies known as artificial intelligence.

Token: In NLP, a token is a unit of text that the system treats as a single item. Tokens can vary in length from a single letter up to several words, and they are numerical representations of the given text. The process of breaking text down into tokens is known as tokenization.

Transfer Learning: An ML technique that leverages knowledge gained from solving one problem or learning on one dataset and applies it to a different but related problem or dataset, reducing the need for extensive training data and computational resources.

Transformer Model: A type of artificial neural network introduced in the paper "Attention is All You Need" by Vaswani et al., from Google Brain, in 2017. It is used primarily in NLP, and the model uses the self-attention mechanism which is essentially a weighted sum of all the input positions. The transformer is a general architecture that excels when the models are pretrained with large amounts of data.

Trigram: See Bigram.

Underfitting: A phenomenon in which an ML model fails to learn the underlying patterns or relationships in the training data, resulting in poor performance on both the training and test data, often due to an overly simplistic model or insufficient training data.

Unstructured Data: Refers to information that doesn't have a predefined model or isn't organized in a predefined manner. This type of data is typically text-heavy but it also includes images and videos. Examples include email, social media posts, videos, audio files, discussion forums, and financial filings. It's more difficult for machines to analyze and interpret compared to structured data.

Unsupervised Learning: A method of ML in which a model is trained using a dataset without ground truth labels, with the goal of discovering patterns, relationships, or structure within the data, often through techniques such as clustering or dimensionality reduction.

Weighted Sum: In the context of neural networks, it's the sum of the input values each multiplied by a corresponding weight, reflecting the importance or contribution of the input to the output.

Wrapper: See Code Wrapper.

Vector: In the context of computing and data science, a vector is a one-dimensional array of numbers. For example, in a three-dimensional space, a vector could be represented as [x, y, z], where x, y, and z are numbers indicating the vector's coordinates in the space. In machine learning, vectors are often used to represent features or parameters in a model. For example, in a model predicting house prices, a vector could be used to represent the features of a house, such as its size, number of rooms, and location.

Zain: A readiness and commitment to learn about and utilize the continually evolving collection of technologies known as artificial intelligence; specifically, a commitment to rapid adoption, pursuit of idiosyncratic data, and a willingness to re-invent processes to maximally leverage these technologies. *From the Arabic word زين (zeyn) which means beauty or grace. It also resembles Zen and Kaizen giving a connotation of both a philosophical approach to reality and continuous improvement and change.*

"If you wish to converse with me, define your terms."

Image created by Midjourney using the description,
"Times Square on New Year's Eve in the style of Monet."