HANDBOOK OF ARTIFICIAL INTELLIGENCE AND BIG DATA APPLICATIONS IN INVESTMENTS

INTRODUCTORY MATERIAL

This book can be found at cfainstitute.org/ai-and-big-data

PREAMBLE

At CFA Institute, our mission is to lead the investment profession globally by promoting the highest standards of ethics, education, and professional excellence for the ultimate benefit of society.

In particular, the closing words—"the ultimate benefit of society"—speak directly to our desire to increase the positive real-world impact of investment management by helping to lead the industry forward and staying abreast of the latest trends, which include leading-edge research into adoption practices of artificial intelligence (AI) and big data.

While many still see AI and big data as a threat, we at CFA Institute have consistently advocated "AI + HI (human intelligence)" as the winning formula for successful financial institutions in the future. Finance professionals bring an important skill set to the table, and embracing the advancement of technology and finding ways to work in harmony with the evolving fintech landscape only lead to better outcomes for our clients and more efficient ways of working. Doing so does require a reimagination of corporate culture and, indeed, the skill sets required among teams to map to this evolution.

As artificial intelligence and big data become increasingly integrated into the investment process, we see this as an opportunity to ensure that those within the industry are well equipped to evaluate and incorporate these factors properly.

Much of this work remains in early stages throughout our industry. What makes our research efforts on these subjects stand out is our dedication to providing the investment professionals' perspectives, which are so often glossed over in favor of the technical aspects of the way the technology functions. But we know the hard work comes when it is time to put these theories into practice. Early adopters' combined experiences are particularly valuable for the industry at large in the implementation of new technology.

Beyond the research we have released to guide the industry on this journey, we are creating active learning opportunities for those in the industry who are interested in mastering the skills required to guide this adoption. Our recently launched professional learning course in data science stands as a prime example in action. We also continue to evaluate and evolve the curriculum of our flagship CFA® Program to take these factors into account.

We extend our deep thanks to our collaborators throughout the industry who have dedicated time and resources to this important work, ensuring the practical nature of this handbook and the integrity of the insights included. It is our hope that this handbook will equip the industry to adopt artificial intelligence and big data practices in a meaningful way for the ultimate benefit of their clients. While the road to adoption can be challenging and complex, it is well worth the commitment. I hope you will find the same to be true on your own journey.

Margaret Franklin, CFA
President and CEO, CFA Institute

FOREWORD

Looking back to the artificial intelligence (AI) work done during the 1980s and 1990s, the new advances in AI and machine learning (ML) over the past decade must surely seem compelling and evolutionary. Breakthroughs in deep learning methodologies and pedagogies are becoming mainstream at a rate never seen before in the marketplace. Many universities have introduced courses in fintech AI (mostly in the computer science, business finance, and mathematics departments), and financial professionals have piled into fintech courses for career development. And companies are feeding this frenzy with capital and recruitment, ranging from ML DevOps to data scientists and scrum teams. While some of the enthusiasm is hype and hyperbole, new discoveries, some of which are quite startling, have made real improvements in how we operate in financial markets.

My first experience in AI was in the early 1990s using ML algorithms in the high-frequency currency markets. Chaos bifurcation, fuzzy logic, genetic algorithms, and perceptrons (early versions of the neural nets) were beginning to become popular in interpreting and trading quotes and prices as well as headline news counts during the late 1980s and early 1990s. But almost as quickly as these approaches gained popularity, markets reverted back to conventional statistical models. There is reason to believe the current wave is more sustainable than previous ones. Development and computing costs have decreased, datasets are getting larger and better defined, ML code libraries have mushroomed, and evolving tech stack architectures are formed ever so quickly to move from development to staging to production. While it can sometimes seem to be a struggle to come to terms with the maze of new wide-ranging enhancements in such a complex space, I believe it is helpful to understand the cost–benefit relationships and how investment and business operating models can benefit from this frontier.

Let us start with language. Developments in natural language processing (NLP) and deep learning are breaking boundaries that seemed difficult to breach before. Word embeddings that can infuse correlation (and thus meanings) across words in vector space have vastly improved ML's natural language understanding (NLU) and NLU capabilities, complemented by an impressive corpus of trained and labeled datasets. This development has allowed the pervasive use of sentiment analysis in handling the huge amount of news and social transmissions of views and opinions. Recent developments of transformer models (e.g., Google's LaMDA) are very close to understanding how we communicate and feel, which perhaps had a big role in considerably attenuating performance alphas for asset management. In addition, the introduction of multilayered architectures in neural networks has provided us a sense of nonlinear formulation of the data features, which has helped propagate the advent of deep learning without the mess of menial feature filtering and wrangling.

But we need to be aware of the challenges to harvesting good results. Sentiment analysis may work well in developed markets that have a free press. Producing similar outcomes in developing countries with heavy news and social media censorship and state propaganda may be difficult. It is not realistic to assume that feature loadings would work with the same effect (without proper adjustments) in such diverse cultural and regime settings.

While many of the breakthroughs are impressive, there is still a long road to reaching (if it was ever possible) the Holy Grail of investments—namely, our ability to produce accurate market forecasts. In this regard, it is useful to explore the differences between financial markets and other disciplines.

First is the importance of the time dimension. Financial analysis differs from most other sectors in the way that data obsolescence occurs, both in speed and manner. This is why we have time-series analysis, a special branch of statistics, that is almost dedicated to the behavior of financial data, as well as dataframes in Python pandas that make it possible to handle the time-dimension property of price data. Success in forecasting market prices depends heavily on time and chronological sequence of one asset vis-à-vis other assets. A consumer planning to purchase an iPhone will have little need to identify exactly when the X13 model price of $1,300 was offered. Conversely, a rational Apple investor would likely have to know the time at which Apple stock traded at, say, $150. Specifically, the decay in the information value of asset prices over time and the conditional information set contained in each asset price data point render the time stamp (of each piece of data) an important property of that data tuple.

This fact leads us to the important work done in GRU (gated recurrent unit), RNN (recurrent neural network), LSTM (long short-term memory), and attention (transformer's encoder-decoder) models—all attempting to link period memory for each piece of data with some form of time embedding. Again, we find that the limitations of these various approaches leave the results wanting. Bayesian neural networks offer promising advantages on top of the time-sequence protocol. Possible solutions may borrow from how human brains are able to extract a distant memory by similar feature transformations rather than a hard-coded LSTM that has a limited time memory threshold or an attention encoder-decoder transformer that suffers from a missing long-term context. The practical outcome at this stage would be that these networks would be excellent in trading, market making, and arbitrage applications but have less success in long-term investment mandates.

For all the tasks attempted with AI in finance, the elephant in the room must surely be our machine's ability to forecast markets. It is as aspirational as it is frustrating. Investment professionals constantly ask whether markets are efficient, while computer scientists question whether P = NP. If we believe that these two questions are related, then perhaps we may surmise that markets are efficient if and only if P = NP. The massive growth of exchange-traded funds (ETFs) versus active funds in developed capital economies reflects the increasing efficiencies in markets with big, credible data and computational power. Counterpose this perspective against the view within the computer science fraternity that P is not equal to NP (which implies that markets likely are not efficient), and we have potentially contradicting beliefs. And if we give credit to common conjectures that markets become increasingly inefficient as the time series lengthens, then the Grossman–Stiglitz (1980) paradox—that the equilibrium level of market efficiency is driven by the market costs of skills and technology to produce alphas—makes a lot of sense.

Although the P = NP conjecture appears theoretical in nature, it has many implications for the various AI applications and solutions used in capital markets. Deep learning has rendered many non-forecast-based tasks (e.g., clustering or categorization) in finance to be in the complexity class P. But stock market prediction problems are still mostly boxed in the NP class, implying that there is a good algorithm for checking a given possible solution but that there is not necessarily a good algorithm for finding solutions. The evolution of AI and ML has given us better ways of understanding, but not inevitably solving the P versus NP problem. Although current states of AI can still fail to find a polynomial-time algorithm in a specified amount of time, this lens allows us to come closer to understanding whether markets can ever be truly efficient. And although it is possible to build an AI algorithm that searches for poly-time solutions to an NP-hard problem, it may also be that the world of predictions is of such complexity that no perfect solution exists. Notwithstanding this possibility, the practical approach for most stakeholders is to find the best available ML approach (among the prevailing sets used in markets) rather than figuring out the perfect solution.

If the current thinking is to improve AI models with new layers of intelligence embeddings, then it may also be helpful to evaluate the "efficient frontier" of competing models in terms of, say, performance versus complexity (as opposed to returns versus risk). For example, new work in metalearning seeks to design models that can learn new skills or adapt to new environments quickly with few training datasets. How do such dynamic (ever-changing common sense or comprehensive learning) demands affect how we train models?

There is no guarantee that the training set's data-generating process will be the same in the test set in financial markets. Environments and how the markets perceive new information are constantly changing in global markets. Regime changes due to regulations, political economies, policy frameworks, technology, market participants, information quality, and so on, can materially change feature impact in nonstationary ways and render training sets less effective at best and disastrous in worst-case scenarios. Even within the realm of explainable empirical solutions, there have been cases that rely too much on nonrepresentative in-sample data to drive investment decisions. The use of copulas based on stable-period in-sample estimation, which produced overly optimistic ratings of asset-backed securities prior to the Global Financial Crisis (GFC), is a case in point.

Is the data-generation process (DCP) ergodic or stationary? A roulette game is very close to ergodic; the distribution of numbers that have come up in the past is very close to the probability distribution of numbers on the next spin. But a game of blackjack when dealt from a set of cards that are not reshuffled after each hand is much less ergodic. This poses a tricky problem when the training, test, and actual data come from different DCPs. Price discovery and risk–return distributions are created from multidisciplinary viewpoints and opinions in a free marketplace, with varieties of frictions, costs, and regulatory constraints. This richness in contributing features can create nonstationarity and ergodic issues, making the DCP much more difficult to calibrate. We can attribute standard volatilities in prices to noise or lead–lag frictions in price discovery, but the continued presence of black swan events (fooling the best of machines and man) tells us that all is not perfect. Such market characteristics do suggest that current AI states are still better suited to areas of success enjoyed by human-operated models or, at least, AI models with close human supervision. This includes such examples as the higher rate of success in arbitrage trading strategies over forecast-driven approaches.

Finally, it is important to augment this discussion with the ability of AI to produce explainable results. ML projects are usually not standalone but are part of a corporate strategy. Naturally, all competing blueprints bubble up to the C-suite for sponsorship. Unknown risks commonly affiliated with unexplainable methodology can be construed as recipes for black swan–type events and derail even the most promising of programs.

For example, suppose someone offers you two investment plans. Plan A offers a 10-year success rate of 70% and is a strategy based on fundamentals that can be readily understood. Plan B is marketed as a strategy that has a 10-year success rate of 80% but is a black box or cannot be easily understood. Which plan would a typical investor choose? Most rational investors would likely prefer Plan A to Plan B. The weakness of statistical models lies in their limited ability to handle large feature sets (dimensions or variables) compared with ML algorithms. But statistical models possess an advantage that is still unmatched in AI models—that is, the presence of an underlying theory that is based on fundamentals and intrinsic valuation drivers. This reinforcement layer of empirical data verification to human conjectures has always been the hallmark of research. ML may not necessarily be a black box, but there is a tendency of users to assume it is.

Although the Financial Stability Board (2017, p. 34) has indicated that progress in AI shows significant promise, regulators continue to be concerned that the "lack of interpretability will make it even more difficult to determine potential effects beyond the firms' balance sheet, for example during a systemic shock. Notably, many AI and machine learning developed models are being 'trained' in a period of low volatility." Reasons for such caution are not hard to find. The 1987 Black Monday stock crash was attributed by many to prevalent option-based program trading at that time. And in the subprime crisis of 2007, there appears to be evidence that the asymptotic independence modeling of the Gaussian copula used to price subprime default likelihood was based on a limited time dataset (stable cycle) and ignored the conventional wisdom that credit defaults tend to cluster.

Fortunately, there is an increasing body of work that attempts to shed light under the AI hood. An example is how Shapley (1953) values or SHapley Additive exPlanation (SHAP) values for machine learning model features are providing interesting insights in much the same way that *p*-values are used for statistical models. Improving ML transparency in this way not only improves conviction in decision making, but it is also helping to build new research to further our understanding of deep learning.

To conclude, it bears mentioning that the current wave of AI does not relate only to new methodologies, big data, and computational power. The additional dimension this time is the democratization of AI, with the wide involvement of professionals, researchers, educators, and policymakers. The massive allocation of resources in this field is testimony to the belief that the benefits are real. However, to make it sustainable beyond the hype of yet another wave, we need to take stock of the promise, manage our expectations, and continue to embrace differing approaches. We are proud to be part of this effort to support and disseminate the body of knowledge through this handbook.

Aaron Low, PhD, CFA
Chair, Board of Trustees,
CFA Institute Research Foundation

References

Financial Stability Board. 2017. "Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications" (1 November). www.fsb.org/wp-content/uploads/P011117.pdf.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70 (3): 393–408.

Shapley, Lloyd S. 1953. "A Value for n-Person Games." *Contributions to the Theory of Games* 2 (28): 307–18.

NOTES FROM THE REVIEWERS

The integration of AI applications into the investment process is gradually becoming part of modern asset management as AI innovation happens at a rapid pace. I was delighted to recommend contributors to and review this handbook, a continuation of the work of CFA Institute in the area. Over the years, I have had the opportunity to engage with innovators, subject matter experts, and practitioners focused on the integration of AI in the investment process and participated in spirited discussions as the risks, challenges, and realities of applications and, importantly, adoption, came to light and we moved from an AI revolution to evolution.

While some think of AI and big data as a threat, there is an opportunity for AI and human intelligence to evolve together to drive better investor outcomes. This requires putting theory and potential use cases into practice. This handbook provides applications currently used by investment practitioners, who share insights, successes, and challenges. Designed and curated by Larry Cao, this handbook is an invaluable reference for investment professionals on the road to adoption as AI innovation transforms our industry.

Carole K. Crawford, CFA
Managing Director, Americas and Global Events
CFA Institute

We are in an era that is both blessed and cursed by data. The pace of our accumulating data is extraordinary, and that makes distilling insights from data harder than ever. When thinking about the investment process, I emphasize the integrity and quality of the decision-making process. However, both are difficult to achieve in real life.

Most of us are trained at school with a generalized foundation of knowledge (broad based but shallow), and then we step into the real world during a certain period of time (experience is linked to a certain period). But the investment industry advocates specialization (depth of knowledge). Take myself as an example: I am an engineer by training, and I joined the investment industry just before the GFC and have been trained as a macro strategist. This means there is a built-in bias in the way I digest information (macro, not micro) and respond to the environment (strong risk aversion). Data science is a perfect tool to help mitigate these biases. State Super has been incorporating data science to support our decision-making process in the last seven years, and the output has been value accretive.

When I was asked to review this book, I was super excited, partly because I wholeheartedly believe in the importance of topics included in this handbook and partly because we need such a publication in our industry. I hope everyone enjoys this book and makes it a companion on their data science journey in the following years.

Charles Wu, CFA
Chief Investment Officer,
State Super SAS Trustee Corporation
President, CFA Society Sydney

It was very gratifying to have been asked to review this handbook. In my current role, I have had the opportunity to interact with numerous experts in the field of AI/ML investing and analysis and have gained an informed sense of the major breakthroughs and the extensive progress that have been made across the space in recent years. Having said that, many of the methods described in this handbook are still in the early stages of development. This seems to me, however, to be a strength; the reader takes away insights from practitioners who openly share their experiences in a dynamically changing environment, where trial and error remain part of the DNA.

Through engaging and straightforward writing, each chapter provides perspectives and often state-of-the-art methodologies on a different aspect of AI/ML in investing. The content is useful both for professional practitioners who are already working in the field and for those who are wishing simply to understand more about how these applications are rapidly changing the landscape of finance and investment management.

The chapters have been carefully selected and edited by Larry Cao, who can again be congratulated on pulling together a wide breadth of pertinent topics and authors and combining them with great skill.

Susan Spinner, CFA
CEO, CFA Society Germany

ABOUT THE CONTRIBUTORS

Mike Chen, PhD, Robeco

Andrew Chin, AllianceBernstein

Kai Cui, PhD, Neuberger Berman

Dan Duggan, PhD, Goldman Sachs Global Investment Research

Yuyu Fan, AllianceBernstein

Che Guan, AllianceBernstein

Stefan Jansen, CFA, Applied AI

Peer Nagy, Man Group

Jochen Papenbrock, Doctorate, NVIDIA

James Powrie, PhD, Man Group

Mikhail Samonov, CFA, Two Centuries Investments and CultureLine.ai

Tal Sansani, CFA, Off-Script Systems and CultureLine.ai

Jonathan Shahrabani, Neuberger Berman

Erin Stanton, Virtu Financial

Ingrid Tierens, PhD, CFA, Goldman Sachs Global Investment Research

Huib Vaessen, APG Asset Management

Xu Liang, PhD, Ping An OneConnect

Weili Zhou, CFA, Robeco

Stefan Zohren, PhD, Man Group

INTRODUCTION

Larry Cao, CFA
Senior Director, Research, CFA Institute

Over the last few years, in our interactions with readers of our artificial intelligence (AI) publications and audiences for our in-person and virtual presentations around the world, we have been repeatedly asked how they should begin their journey of AI adoption. This book is our attempt at providing initial answers to their questions.

Who Should Read This Handbook?

We have produced this handbook with the following target audience and objectives in mind:

- Help C-suite executives and board members with responsibility to develop and execute their firms' AI and big data strategies to get an overview of the subject and set strategic directions.
- Help knowledge engineers leading AI and big data projects at investment firms and their teams select projects.
- Help investment and technology professionals who work on T-shaped teams understand project backgrounds and implementation details.
- Help regulators stay abreast of industry developments so as to facilitate their policy development efforts.
- Help students (and the professors who are teaching them) prepare for a career on future investment teams or at regulatory agencies.

In short, if you are embarking on the journey of AI and big data adoption in finance, this book is written for you.

The AI Adoption Journey in Finance and You

There is a well-known five-stage model of the buyers' decision process for new products (Kotler and Armstrong 1998). As we studied the benefits, hurdles, and ways to overcome these challenges of AI adoption over the last few years, we came to realize that the AI adoption journeys of financial institutions and finance professionals have largely followed patterns that mirror this five-stage process. Here is a summary of our findings, which culminated in this handbook.

The five stages are awareness, interest, evaluation, trial, and adoption (see **Exhibit 1**). It is important to note that buyers have different needs at each stage. For example, at the awareness stage, they lack information about the new product, which they seek at the interest stage. At the evaluation stage, buyers look for more specific proof that the new product makes sense, and if so, they will move to the next step, trial, where they experiment with the new product. If all goes well, they will finally reach the adoption stage and start using the new product regularly.

Awareness

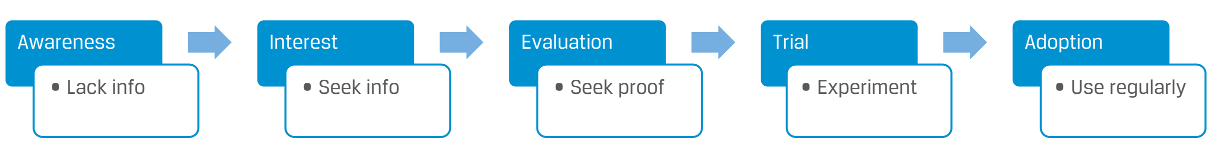
From 2015 to 2017, AlphaGo went from beating a human professional player at the board game Go for the first time to beating the best human Go player. The historic events brought artificial intelligence from computer labs into the public domain, quickly becoming the subject of conversation at dinner tables, at cocktail parties, and in board rooms.

Up to that point, main street and the mainstream financial services industry had focused their attention on what we later referred to as "early-stage fintech"—that is, peer-to-peer lending, mobile payments, and robo-advice. Our investigation led us to refute that these new businesses were disrupting the financial services industry, simply because they prospered where customers were un(der)-banked and fared less well where financial services were entrenched.

I also wrote in the financial press in 2016 that we believed winners would follow the model "Fin + Tech" (Cao 2016)—that is, collaboration between powerful financial institutions and powerful tech companies. The winner-takes-all model commonly seen in the internet business will not extend quite as easily to finance.

The first publication in our series, "FinTech 2017: China, Asia, and Beyond" (CFA Institute 2017), included a contributor's

Exhibit 1. The Five Stages of the Buyers' Decision Process for New Products



article highlighting that the ABCs of fintech (artificial intelligence, big data, and cloud computing) will be the driving force for change in the financial services industry. Our audience's reception, however, was lukewarm in the spring of 2017.

Interest

Google Trends showed that worldwide searches for "artificial intelligence" continuously rose for several years through 2018. This finding is consistent with our experience. We received more media requests on AI and were quoted in an Ignites article in the summer of 2017 on our positive views regarding the collaboration between Microsoft and China Asset Management, one of the top mutual fund firms in China.

Six months later, we invited Tang Xiaodong,[1] CEO of China Asset Management; Eric Chang, vice president of Microsoft Research Institute Asia; and senior executives from other sectors in finance to speak at the AI and the Future of Financial Services Forum, a major event we organized in Beijing, which was also livestreamed to satellite venues in several other financial centers in the region. It attracted a record number of participants across the region—either joining in person in Beijing or at one of the satellite venues.

In his fascinating presentation on machine learning (ML) and neural networks, Chang highlighted what Microsoft believed was going to be the winning formula of AI adoption: "AI + HI (human intelligence)," which we readily adopted because we considered it an extension of our "Fin + Tech" philosophy in the age of AI.

Tang gave an informative talk on AI's applications in asset management. Other speakers rounded out the forum with insights from their unique vantage points. The event received the highest ratings from participants, reflecting both the audience's passion for the subject and their satisfaction with our expert speakers.

My articles on the subject, from the concepts of AI, machine learning, and deep learning (Cao 2018a) to their applications in finance (Cao 2018b), published, respectively, in February and March 2018 on the *Enterprising Investor* blog were similarly popular with readers, such that the former became the blog's top ranked article of February 2018 (McCaffrey 2018a); these two articles also ranked among the top 10, for the year (McCaffrey 2018b). These articles and more from our contributors formed the core of our "FinTech 2018" report (CFA Institute 2018).

Evaluation

In late 2018, I was invited to speak at two closed-door events, one attended by technology executives at global financial institutions and the other with researchers from leading central banks and universities around the world. My fellow speakers at both events shared one frustration: They were having a hard time distilling insights from big data!

The participant demographic at both events was certainly representative of the leaders in their respective domains, which seemed to me a clear sign that AI adoption in financial services had reached what Gartner refers to as "the trough of disillusionment" in the Hype Cycle model,[2] where "interest wanes as experiments and implementations fail to deliver." In the simpler terms of the five-stage model, our audience reached the evaluation stage and needed proof that AI works before they could move forward.

We believed the best way to provide that proof would be a collection of case studies from successful early adopters who had put AI and big data to productive use in finance. To help readers relate, we selected cases based on two criteria: geography—Americas; Asia Pacific (APAC); and Europe, Middle East, and Africa (EMEA)—and line of work—stocks, bonds, asset allocation, and hedge funds. These elements formed the basic construct of the "AI Pioneers in Investment Management" report (CFA Institute 2019).

We tested the idea while speaking at the forum of the CFA Institute Annual Conference in 2019,[3] and judging from the audience's reaction and questions, they were clearly looking for concrete examples, exactly the type that we were working on, rather than a continuation of the more generic type of information we provided in 2018.

A CFA Institute Practice Analysis survey conducted in the spring of 2019 showed that adoption rates for AI and big data techniques in the industry remained low (CFA Institute 2019, pp. 8–13). We shared our perspectives in the report on the challenges the industry faced and provided our recommendation. The report was cited in the *Financial Times* and the *Harvard Business Review* and by Reuters and researchers at Deloitte and Morningstar, among others.

Trial

The day after the report was published, I spoke at an industry conference in Toronto.[4] And then I went to Singapore, before going on to speak at more than a dozen financial

---

[1]He left China Asset Management in 2019 and now leads BlackRock's China efforts (BlackRock 2019).

[2]See the "Gartner Hype Cycle" webpage at www.gartner.com/en/research/methodologies/gartner-hype-cycle.

[3]See my comments in the video "AI and Big Data in the Future of Investment Management" at https://cfainstitute.gallery.video/annual19/detail/video/6036457979001/ai-and-big-data-in-the-future-of-investment-management.

[4]Toronto, and broadly Canada, is of course also where this latest wave of AI development started.

centers in APAC and EMEA, providing us with many additional touch points with the industry and finance professionals to get a read on their needs.

The feedback we got from our readers and audiences was overwhelmingly consistent with the model prediction: They wanted to experiment with AI and big data and asked for help on where to begin.[5] Of course, every firm and every professional are unique. Providing a one-size-fits-all answer is not only challenging but also likely to be counterproductive. That said, we believe we can provide a framework, based on our own analysis and field research into cases that were successful and, even more importantly, those that were less fortunate.

The first step in the framework is building an AI-enabling organization, or T-shaped teams. We developed the concept of a T-shaped team specifically to highlight the need for communication and strategic leadership in the context of AI adoption in financial services (CFA Institute 2021). A T-shaped team comprises three distinct functions: the traditional investment function, a technology (data science) function, and an innovation function. The background and training of finance and data science professionals are so different that it makes knowledge engineers in the innovation function who can provide both communication support and strategic leadership especially important for AI adoption to succeed.

The second step in the framework is to implement the AI strategy developed by the leadership. For this step, we believed it would be most helpful to provide both executives at financial institutions looking to embark on the AI adoption journey and professionals at these institutions a menu of options they could pick and choose from that would be most appropriate for their strategy, product lineup, and staff's level of expertise in and acceptance of AI adoption. This is where this handbook comes in.

There is a plethora of books published about machine learning for investment and finance. Yet they tend to be written for quants or programmers with prior machine learning knowledge and focus on the technical aspects of implementation. The attention tends to be 80/20 between technology and finance.

This handbook is different. We strive to provide a reference book on AI and big data applications in investments written from the finance industry perspective and offer details on solutions that are in production and effectively field-tested every day.[6]

Consistent with this objective, we have encouraged contributors to include snippets of coding where they deem appropriate. We hope the more tech-savvy readers among us will find it helpful, and that the rest of us will be motivated to catch up over time.

We have summarized from our vantage point the industry's progress on the AI adoption journey and our efforts to support it (see **Exhibit 2**). What is next? The most natural scenario is that we will support the next stage of the adoption journey in much the same way as we do in other, more established areas of investment. Do let us know what you think and how we can best be of service.

Exhibit 2. The Five Stages of AI Adoption and Accompanying Research

[5]Another popular area of interest that emerged at the time was AI ethics. For additional details, see CFA Institute (2022).

[6]Many may consider "field-tested" an aspirational goal in AI and big data applications today, probably rightly so. Our operating definition of the term is that the type of solution is in production at different firms in the process of managing significant sums of money.

How Is the Handbook Organized?

In line with our understanding of the active areas of AI and big data adoption in investments (and, more broadly, finance), we have organized the handbook along four streams, each with a few chapters written by practitioners who are active in that area. Representative of the ecosystem, most of our contributors hail from investment firms, with the remaining coming from investment banks, other financial institutions, and service vendors.

Part I provides an overview of the current state of AI and big data adoption in investments. In Chapter 1, Mike Chen and Weili Zhou share their perspectives on challenges and potential pitfalls in applying ML in finance and provide an overview of the broad areas of applications. In Chapter 2, Ingrid Tierens and Dan Duggan discuss their experiences and takeaways in working with alternative data, in addition to some use cases. Chapter 3, by Kai Cui and Jonathan Shahrabani, reviews how data science can inform long-term investment decisions of active fundamental investors.

Part II covers natural language understanding, processing, and generation. Natural language processing (NLP) is a branch of AI studies that dates to the 1950s, although the latest boom is often attributed to the introduction of ML into the field since the 1980s. Its applications in investments also go back well over a decade (see, e.g., Antweiler and Frank 2004), although significant strides have been made since then, which our contributors for Chapters 4–6 focus on.

In Chapter 4, Andrew Chin, Yuyu Fan, and Che Guan provide an overview of the broad applications of natural language processing in investments today, ranging from well-known areas, such as sentiment analysis and earnings transcript analysis, to more recent applications using corporate filings in compliance and, in sales, to gain client insights from public information. In Chapter 5, Stefan Jansen discusses from a technical perspective the evolution of natural language understanding and trends in its applications. Chapter 6, by Tal Sansani and Mikhail Samonov rounds out this stream with an aspirational discussion on applying NLP in environmental, social, and governance analysis.

Part III covers trading. This is an active area of AI and big data activities, although the percentage of trades executed with AI and big data techniques remains small. Our contributors from different parts of the ecosystem shared their perspectives. Chapter 7, by Erin Stanton, highlights a few use cases at various stages of the trading process that, in aggregate, help readers piece together a picture of ML and big data application in the trading landscape. In Chapter 8, Peer Nagy, James Powrie, and Stefan Zohren of Man Group zoom in on limit order books as the microstructure data source and explain in detail how ML models can help predict spreads, trade volume, volatility, and more to inform trading decisions.

In addition to the areas mentioned thus far, there are obviously still many more AI and big data applications that are being developed and applied in investments and finance. Part IV addresses development in three of these areas.

Customer service (call centers) is one of the most tried-and-true applications for AI that has been widely applied across industries. In Chapter 9, Xu Liang of Ping An OneConnect offers an overview of financial institutions' use of AI in customer service, providing many enlightening business applications and discussing the underlying technology, such as voice, NLP, and knowledge graphs.

The current AI boom is largely driven by progress in three areas—algorithms, big data, and computing power—where computing power's contribution is often understated. In Chapter 10, Jochen Papenbrock of NVIDIA argues, supported by a number of recent cases in investments, that there is increasingly a need for accelerated computing.

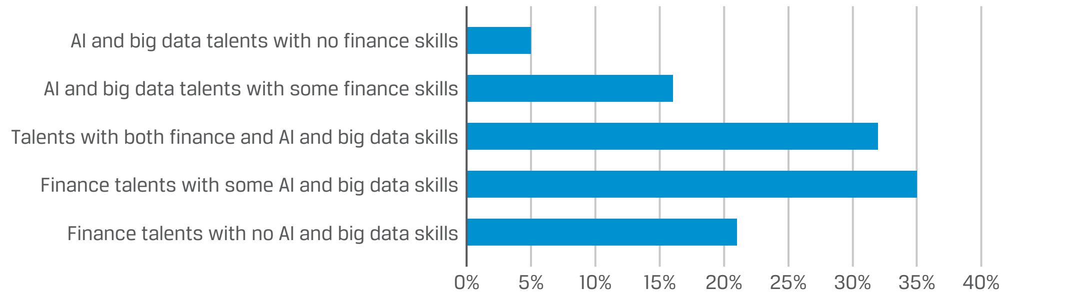
AI is a fluid term. The scope of our AI and big data efforts, including this handbook, focuses on machine learning (and its various branches)—NLP, computer vision, and unstructured data. Symbolic AI is a branch of AI research that was popular before the latest AI boom, brought on by nonsymbolic AI studies, such as machine learning. In the last chapter of this handbook, Huib Vaessen shares how an automated real estate portfolio management solution based on symbolic AI can help an investment team.

The results of a CFA Institute global survey conducted when this handbook was completed support our positioning and topic selection (see **Exhibit 3**). To the question of which talents will be in most demand, 35% of respondents answered finance/investment talents with some AI and big data training and 32% answered both finance/investment and AI/big data skills. This handbook has indeed been written to help finance/investment professionals move into those two buckets.

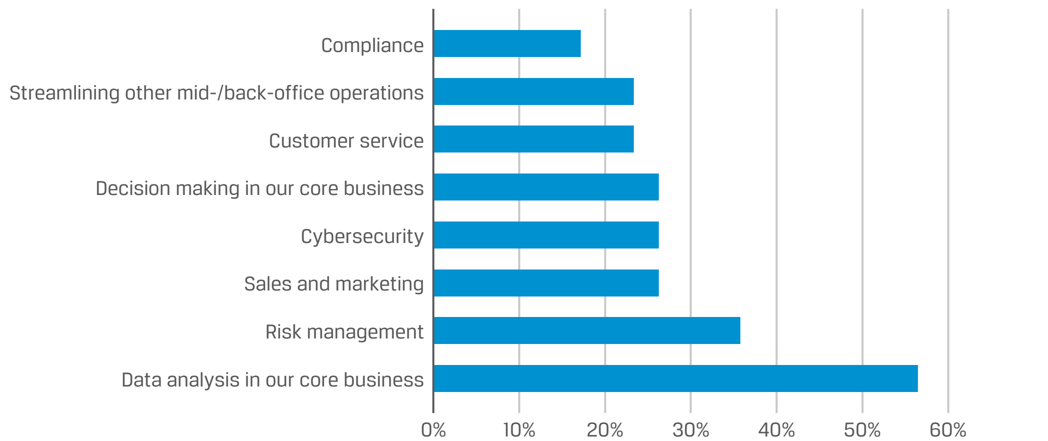
The picture is similar in terms of topic selection. For example, 56% of the respondents are using AI and big data in data analysis, a topic that we hope this handbook has thoroughly covered throughout, and about a quarter of the respondents are using AI and big data techniques in areas outside the core investment business, such as sales and marketing or customer service, which are also well covered in the handbook. Chapter 9 focuses entirely on customer service. Chapter 4 covers applications in sales and marketing in addition to investment applications, which is partly why we allocated more space to the chapter.

Exhibit 3. Responses to CFA Institute Global Survey Questions

A. Over the next two years, which type of talents is your organization primarily looking to hire more? (select up to two)



B. What business activities has your organization been routinely applying AI and big data to? (select all that apply)



As we continue our research efforts in AI and big data applications, please feel free to drop us a note and share with us your successes or needs. We will strive to serve as your knowledge engineers, building bridges between thought leaders and practitioners.

Acknowledgments

Putting together such a handbook of unique positioning for the first time was not an easy task. My gratitude goes to all who have supported our efforts.

First, I have been fortunate to work with some of the key industry practitioners putting AI and big data to work in investments and finance. Despite the tremendous demand on their time, they have all given the project much attention, often patiently working with us through multiple iterations to make sure that the chapters collectively form a coherent chorus.

My deep appreciation also goes to Aaron Low, CFA, Carole Crawford, CFA, Charles Wu, CFA, and Susan Spinner, CFA, for reading through the handbook with me from its infancy in the spring of 2022 through the final draft in the fall. Their expertise, judgment, and network have gone a long way in raising the bar for this project.

Finally, credit goes to the leadership and the project team at CFA Institute. Without their steadfast support, this handbook would have never been able to get to where it is today.

References

Antweiler, W., and M. Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance* 59: 1259–94.

Bartram, Söhnke M., Jürgen Branke, and Mehrshad Motahari. 2020. "Artificial Intelligence in Asset Management." Literature review, CFA Institute Research Foundation (28 August). www.cfainstitute.org/research/foundation/2020/rflr-artificial-intelligence-in-asset-management.

BlackRock. 2019. "BlackRock Appoints Tony Tang Head of China" (17 April). www.blackrock.com/corporate/newsroom/press-releases/article/corporate-one/press-releases/blackrock-appoints-tony-tang-head-of-china.

Cao, Larry. 2016. "FinTech and the Future of Financial Services." *Hong Kong Economic Journal* (26 May). http://startupbeat.hkej.com/?p=29681.

Cao, Larry. 2018a. "Artificial Intelligence, Machine Learning, and Deep Learning: A Primer." *Enterprising Investor* (blog; 13 February). https://blogs.cfainstitute.org/investor/2018/02/13/artificial-intelligence-machine-learning-and-deep-learning-in-investment-management-a-primer/.

Cao, Larry. 2018b. "Portfolio Managers, Artificial Intelligence Is Coming for Your Jobs." *Enterprising Investor* (blog; 9 March). https://blogs.cfainstitute.org/investor/2018/03/09/portfolio-managers-artificial-intelligence-is-coming-for-your-jobs/.

CFA Institute. 2019. "AI Pioneers in Investment Management." www.cfainstitute.org/-/media/documents/survey/AI-Pioneers-in-Investment-Management.pdf.

CFA Institute. 2022. "Ethics and Artificial Intelligence in Investment Management: A Framework for Investment Professionals." www.cfainstitute.org/-/media/documents/article/industry-research/Ethics-and-Artificial-Intelligence-in-Investment-Management_Online.pdf.

CFA Institute. 2017. "FinTech 2017: China, Asia, and Beyond" (May). https://hub.ipe.com/download?ac=72587.

CFA Institute. 2018. "FinTech 2018: The Asia Pacific Edition" (September). www.cfainstitute.org/-/media/documents/book/industry-research/fintech-booklet.pdf.

CFA Institute. 2021. "T-Shaped Teams: Organizing to Adopt AI and Big Data at Investment Firms." www.cfainstitute.org/-/media/documents/article/industry-research/t-shaped-teams.pdf.

Kotler, Philip, and Gary Armstrong. 1998. *Principles of Marketing*. Hoboken, NJ: Prentice Hall.

McCaffrey, Paul. 2018a. "Top Five Articles from February: Body Language, Data Science, Artificial Intelligence." *Enterprising Investor* (blog; 28 February). https://blogs.cfainstitute.org/investor/2018/02/28/top-five-articles-from-february-body-language-data-science-artificial-intelligence/.

McCaffrey, Paul. 2018b. "Top 10 Posts from 2018: Chart Crimes, Political Divides, AI, and the Skill Ratio." *Enterprising Investor* (blog; 27 December). https://blogs.cfainstitute.org/investor/2018/12/27/top-10-posts-from-2018-chart-crimes-political-divides-ai-and-the-skill-ratio/.