



## Opening the Black Box: Why AI Needs to Be Transparent (2 of 5)

Just as ancient man attributed natural phenomena to the gods without question, so too do many modern humans blindly trust in the power of AI. However, for AI to be a truly effective tool in finance, transparency about its inner workings should be a natural by-product of the integrity of its design and effectiveness.<sup>1</sup>

*"I have bought this wonderful machine—a computer. Now I am rather an authority on gods, so I identified the machine—it seems to me to be an Old Testament god with a lot of rules and no mercy."*

-Joseph Campbell, The Power of Myth

The idea of the computer as a god has existed almost as long as the computer itself. In 1954, science fiction writer Isaac Asimov turned the computer into a literal god in his short story, "[The Last Question](#)." In this story, generation after generation of humans beseech a galactic computer to stave off the end of the universe, but only after the universe dies does this computer begin the act of recreation with the words, "LET THERE BE LIGHT!"

Even working with the primitive computers of the 1950's, Asimov understood how complex the machines could become:

*"Multivac (the computer) was self-adjusting and self-correcting. It had to be, for nothing human could adjust and correct it quickly enough or even adequately enough. So Adell and Lupov attended the monstrous giant only lightly and superficially, yet as well as any men could. They fed it data, adjusted questions to its needs and translated the answers that were issued."*

Anyone working in AI today doubtless recognizes themselves in Adell and Lupov. They feed data in, try to make sense of the results, and do their best to maintain a black box of a system no one truly understands.

### How We Got Here

The development of artificial intelligence as a black box may have been predictable, but it was far from inevitable. The lack of transparency came about due to the fact that the people developing the backbone of AI systems and the ones tasked with deploying those systems have little in common.

In fact, these disparate folks tend not to work under the same roof. While a large percentage of companies claim to be [developing AI technologies](#), the truth is that most of them are just repurposing others' research. The bulk of AI programs are [being developed in universities](#) and by a handful of tech giants like Google (GOOGL), Microsoft (MSFT), and Facebook (FB).

The researchers at these institutions create AI algorithms for a variety of purposes. Developers at other companies discover that these algorithms can do 80% of what they need and hack together modifications to try to get the other 20%.

The end result is a program that no one fully understands. The initial researchers don't know how the AI is being used. The developers don't understand how the underlying code, that they get from some another party, works (or, more often, does not work). The actual users couldn't get a clear picture of how the AI makes decisions even if they tried – which they usually don't.

Another part of the problem is simply the newness of AI. Researchers have been so focused on teaching computers to "think" in a way that produces useful results that they haven't spared as much thought on making those thought processes transparent. Further, given the lack of success in getting AI to produce useful results, there are not a lot of inner workings to brag about. Why show off something that does not work so well?

Slowly but surely, the technological community is beginning to question the black box model. After all, AI exists because it's supposed to outstrip human intelligence at specific tasks. If the AI's thought process is superior,

---

<sup>1</sup> Harvard Business School features the powerful impact of our research automation technology in the case [New Constructs: Disrupting Fundamental Analysis with Robo-Analysts](#).



shouldn't its explanation of that thought process be equally sophisticated? As philosopher and cognitive scientist Daniel Dennett told the [MIT Technology Review](#):

*"If it can't do better than us at explaining what it's doing, then don't trust it."*

### The Problems of the Black Box

Dennett identifies the most pressing reason why AI needs to become more transparent. As long as AI stays a black box, many will be hesitant to trust it. When it comes to handling people's money, this lack of trust becomes even more pronounced.

The common rebuttal to this concern has been that the superior returns of AI-driven investment decisions will win skeptics over. Putting aside the fact that AI hedge funds [haven't yet succeeded in delivering outperformance](#), there's reason to believe that returns alone won't be enough to attract investor capital.

Research shows that the best performing asset managers [don't necessarily attract the most new capital](#). The most successful funds at attracting new capital do so by building investor trust through transparency and investor education. Many investors prefer solid returns they understand over great returns they don't.

These investors are right not to trust black boxes. Even if AI succeeds in identifying unseen patterns to deliver superior returns, that success will be only temporary without transparency. If no one understands how the machine works, they won't be able to figure out when it stops working correctly, much less how to fix it.

Right now, many companies that implement AI have to do so through a process of trial and error. If the machine fails to correctly process a data set, they just have to keep (often blindly) tweaking different parameters up and down until it spits out the right results.

Trial and error might work in the "move fast and break things" world of Silicon Valley, but it's not going to fly when investors' money is on the line.

### How to Open the Black Box

As AI becomes more ingrained in every part of our lives, people are making a concerted effort to fix the technology's transparency problem. The AI Now Institute recently issued a [set of recommendations](#) for the AI industry: goal number 1 was that core public agencies should no longer use black box AI systems.

On the finance side, the EU's [General Data Protection Regulation](#) could impose punishments on companies that rely on black box algorithms. While the Fiduciary Rule doesn't touch on AI specifically, you have to believe that black box technologies won't be enough to fulfill the [Duty of Care](#). What regulator would accept "the computer told me to do it" as the rationale for an investment decision?

In order to fulfill the goal of transparency, researchers have put a lot of effort into teaching AI to explain itself. Researchers at MIT successfully built a neural network with two different modules that could successfully make simple predictions and [highlight its reasoning](#). Developments such as these prove that the black box can be opened.

In order to end the black box, financial companies working on AI should take a few concrete steps:

1. Take a holistic approach to development where those with tech expertise and those with finance expertise work together on every stage of the project.
2. Make explicability a key goal. An AI that can't explain its reasoning is no true AI.
3. Commit to auditability. As much as possible, make the data inputs and outputs accessible so that others can verify the machine's "thought" processes.

These steps seem obvious. Point 1 is clearly the best strategy for developing any technology, and 2 and 3 are not only good in principle, they're good for business. If you've done the work to build great technology, why wouldn't you want to show clients all the work you've done? It's easier to build trust in products clients can see and understand.

We've seen progress on point 1. Certain financial firms are making an effort to bring more tech knowledge in-house, as banks like JPMorgan Chase (JPM) are [going head-to-head with tech giants](#) to hire AI researchers and developers.



The second point has been more of a mixed bag. Credit card companies like Capital One (COF) have [dedicated research teams](#) trying to make their computer techniques more explainable. On the other hand, many asset managers still seem content to rely on black box technology.

The final point can be the most difficult, but it's also incredibly important. Financial firms often use the sensitivity of their data as an excuse to keep it hidden even when they could potentially anonymize and share it. Without the availability of this data, it becomes much harder for investors to trust in AI. No matter how sophisticated the technology is, it will be useless without the right data. Garbage in, garbage out.

### **Speaking from Experience in Building a Transparent Box**

In the development of our machine learning and AI technology, we've found that commitment to a holistic approach, explicability and auditability pays many dividends. It boosts the trust of our clients in our work (e.g. our [click-through filings](#) service), and augments our ability to write better code and improve the AI.

On the holistic approach: since our inception, we've rallied around the benefits of combining different kinds of expertise. We eschew the idea of siloed teams. Empirical evidence proves the benefits of working in teams built on [different experiences and skill sets](#). Open communication between analysts and programmers allows us to anticipate and address problems in advance while also reviewing our work with a critical eye. We see great competitive advantage in our ability to combine technological and finance expertise.

On explicability: by clearly documenting our code and using consistent formatting that everyone on the team understands, we're able to work efficiently with a larger team. We avoid any reliance on one person going down a rabbit hole that no one else understands. Explicability keeps all our programmers on the same page and provides excellent training material for bringing new programmers up to speed. It also ensures there is one version of the truth about how our code works.

In addition, all project requests from our financial analysts are required to conform to specific standards so that both parties know exactly what is expected of the other. We derived these standards over 15 years of successes and failures in machine learning and AI development. One of the most important benefits of these standards is the improved transparency and communication they create for the different minds working on projects. Plus, they ensure there is one version of the truth about how the teams will execute the project and test its success.

On transparency: having a better understanding of how code changes affect output makes communicating about how code does or does not work easier. When the team is able to directly and discretely measure the output or impact of code changes, then it can directly measure the efficacy of the code and communicate how it may need to change. Transparency enables everyone to be more aware and, therefore, smarter about the process and the results. Running all of our data through financial models that we publish to clients unifies the firm around a discrete set of outputs that we can manage and measure.

Transparency boosts external communications as well. We show the source filing data and all the calculations our [Robo-Analyst](#) uses to build our models and determine investment ratings. We want clients to understand how much work our machines do and be able to verify thought processes. There's no reason for us (or any company) to hide how much work, planning and sophistication goes into our technology.

This article is the second in a five-part series on the role of AI in finance. The first, "[Cutting Through the Smoke and Mirrors of AI on Wall Street](#)" highlighted ways AI is actually being and not being used on Wall Street. In the next article, we will dig deeper into the challenges facing AI and how they can be overcome, while the last two articles will show how AI can lead to significant benefits for both financial firms and their customers.

*This article originally published on [January 22, 2018](#).*

*Disclosure: David Trainer and Sam McBride receive no compensation to write about any specific stock, sector, style, or theme.*

*Follow us on [Twitter](#), [Facebook](#), [LinkedIn](#), and [StockTwits](#) for real-time alerts on all our research.*



## ***New Constructs® - Research to Fulfill the Fiduciary Duty of Care***

---

Ratings & screeners on 3000 stocks, 450 ETFs and 7000 mutual funds help you make prudent investment decisions.

New Constructs leverages the latest in machine learning to analyze structured and unstructured financial data with unrivaled speed and accuracy. The firm's forensic accounting experts work alongside engineers to develop proprietary NLP libraries and financial models. Our investment ratings are based on the best fundamental data in the business for stocks, ETFs and mutual funds. Clients include many of the top hedge funds, mutual funds and wealth management firms. David Trainer, the firm's CEO, is regularly featured in the media as a thought leader on the fiduciary duty of care, earnings quality, valuation and investment strategy.

### ***To fulfill the Duty of Care, research should be:***

1. **Comprehensive** - All relevant publicly-available (e.g. 10-Ks and 10-Qs) information has been diligently reviewed, including footnotes and the management discussion & analysis (MD&A).
2. **Un-conflicted** - Clients deserve unbiased research.
3. **Transparent** - Advisors should be able to show how the analysis was performed and the data behind it.
4. **Relevant** - Empirical evidence must provide [tangible, quantifiable correlation](#) to stock, ETF or mutual fund performance.

### ***Value Investing 2.0: Diligence Matters: Technology is Key to Value Investing With Scale***

Accounting data is only the beginning of fundamental research. It must be translated into economic earnings to truly understand profitability and valuation. This translation requires deep analysis of footnotes and the MD&A, a process that our [robo-analyst technology](#) empowers us to perform for thousands of stocks, ETFs and mutual funds.



## ***DISCLOSURES***

New Constructs®, LLC (together with any subsidiaries and/or affiliates, "New Constructs") is an independent organization with no management ties to the companies it covers. None of the members of New Constructs' management team or the management team of any New Constructs' affiliate holds a seat on the Board of Directors of any of the companies New Constructs covers. New Constructs does not perform any investment or merchant banking functions and does not operate a trading desk.

New Constructs' Stock Ownership Policy prevents any of its employees or managers from engaging in Insider Trading and restricts any trading whereby an employee may exploit inside information regarding our stock research. In addition, employees and managers of the company are bound by a code of ethics that restricts them from purchasing or selling a security that they know or should have known was under consideration for inclusion in a New Constructs report nor may they purchase or sell a security for the first 15 days after New Constructs issues a report on that security.

## ***DISCLAIMERS***

The information and opinions presented in this report are provided to you for information purposes only and are not to be used or considered as an offer or solicitation of an offer to buy or sell securities or other financial instruments. New Constructs has not taken any steps to ensure that the securities referred to in this report are suitable for any particular investor and nothing in this report constitutes investment, legal, accounting or tax advice. This report includes general information that does not take into account your individual circumstance, financial situation or needs, nor does it represent a personal recommendation to you. The investments or services contained or referred to in this report may not be suitable for you and it is recommended that you consult an independent investment advisor if you are in doubt about any such investments or investment services.

Information and opinions presented in this report have been obtained or derived from sources believed by New Constructs to be reliable, but New Constructs makes no representation as to their accuracy, authority, usefulness, reliability, timeliness or completeness. New Constructs accepts no liability for loss arising from the use of the information presented in this report, and New Constructs makes no warranty as to results that may be obtained from the information presented in this report. Past performance should not be taken as an indication or guarantee of future performance, and no representation or warranty, express or implied, is made regarding future performance. Information and opinions contained in this report reflect a judgment at its original date of publication by New Constructs and are subject to change without notice. New Constructs may have issued, and may in the future issue, other reports that are inconsistent with, and reach different conclusions from, the information presented in this report. Those reports reflect the different assumptions, views and analytical methods of the analysts who prepared them and New Constructs is under no obligation to insure that such other reports are brought to the attention of any recipient of this report.

New Constructs' reports are intended for distribution to its professional and institutional investor customers. Recipients who are not professionals or institutional investor customers of New Constructs should seek the advice of their independent financial advisor prior to making any investment decision or for any necessary explanation of its contents.

This report is not directed to, or intended for distribution to or use by, any person or entity who is a citizen or resident of or located in any locality, state, country or jurisdiction where such distribution, publication, availability or use would be contrary to law or regulation or which would be subject New Constructs to any registration or licensing requirement within such jurisdiction.

This report may provide the addresses of websites. Except to the extent to which the report refers to New Constructs own website material, New Constructs has not reviewed the linked site and takes no responsibility for the content therein. Such address or hyperlink (including addresses or hyperlinks to New Constructs own website material) is provided solely for your convenience and the information and content of the linked site do not in any way form part of this report. Accessing such websites or following such hyperlink through this report shall be at your own risk.

All material in this report is the property of, and under copyright, of New Constructs. None of the contents, nor any copy of it, may be altered in any way, copied, or distributed or transmitted to any other party without the prior express written consent of New Constructs. All trademarks, service marks and logos used in this report are trademarks or service marks or registered trademarks or service marks of New Constructs. Copyright New Constructs, LLC 2003 through the present date. All rights reserved.