AI: PAST, PRESENT, FUTURE

MIKE OLSON, FOUNDER

CLOUDERA

1. WHERE DID ALL THIS HYPE COME FROM?

"Back in the days of Alan Turing, when the fundamental views of computer science were being created, people were already thinking about what it would mean when machines could think."

If you really want to understand the current hype around artificial intelligence, then it is useful to look at history and to understand its origins. Back in the days of Alan Turing, when foundational ideas of computer science were still emerging, people were already thinking about what it would mean when machines could think. In 1950, Turing came up with a very famous thought experiment: the so-called Turing test. The test involves a conversation between two parties, one of whom is a computer and one of whom is human, and an evaluator observing the conversation. If the evaluator cannot tell which party is which, then the computer has passed the Turing test.

The theoretical founders of AI—such as John McCarthy, Marvin Minsky, Ray Solomonoff, John Henry Holland, and others—showed a lot of early enthusiasm about using computers to create thinking systems. Their earliest work started from the premise: We know that people can think, and we have a little bit of an idea about how their brains work; we understand what neurons and synapses are, and we understand that they are connected to one another. So we're going to model in software these neural networks connected to one another with inputs and outputs. They attempted to build software models of the brain, and figured that if they made them big enough, they could create the conditions for intelligence to spontaneously emerge.

In fact, these pioneers built some pretty interesting systems. In 1968 at MIT, Terry Winograd developed a language parsing program called SHRDLU that was capable of parsing natural English. It allowed a user to direct the computer to move around differently shaped blocks on a screen—pick up the green pyramid and set it on the red block. And it worked. There were other kinds of expert systems that were designed as well. Some were built to diagnose diseases or to recognize ocular malfunctions, and provide advice to doctors. These were expert systems, and they were pretty successful.

All of those successes gave early researchers significant hope that the problem of creating true artificial intelligence would be pretty easy to solve.

Recently, Sundar Pichai gave a demonstration of Google Duplex on stage at Google I/O. He played a recording of an automated agent calling up businesses to schedule appointments. The people at the other end of those phone calls certainly had no idea that they were talking to a robot. None whatsoever. Granted, those people didn't know they should be skeptical that they were talking to a human being. Had the conversation observers, the audience at Google I/O, not been informed which was which, that robot would have easily passed the Turing test. On that basis, artificial intelligence exists today.

It's understandable that lots of people are excited about this progress. And yet, we still haven't achieved a computer that can think in the way that people can.

2. BRITTLE SYSTEMS

"Forget the Turing test, because we have now passed it—but nobody believes that those chat bots are actually thinking. They are just implementing very specific models..."

Building intelligent systems turns out to be a much harder problem than the optimism of early computer scientists had suggested. Systems to parse natural language and respond to it in specific ways, such as SHRDLU, could be built—and they gave impressive early results—but they were not generalizable.

If you specialized a system in one type of information, it couldn't then also analyze information on a separate subject that the system hadn't been specifically trained on. All of the knowledge of these computers had to be hard-coded—and so they were brittle. As soon as you stepped 2 or 3 degrees off of the specific problem that the system had been built to solve, it didn't work at all. And it was exactly as expensive to go build a new system that was capable of doing tasks those 2 or 3 degrees off as it had been to build the first system. We had nothing general-purpose.

There are certain algorithmic techniques that we use in a wide variety of applications and use cases. They are powerful, but they are still very brittle. They are still custom-tuned for a narrow set of data—and therefore a narrow question—because at their creation they were trained on particular data in order to answer a particular question.

If you teach an algorithm to spot credit card fraud, it will not be able to answer you when you try to order takeout pizza. It's not even of any use in patient readmission prediction. Human beings are able to generalize and draw analogies—bad guys are pretty uncommon but very clever, and disease is pretty uncommon, but in some sense very clever—but we have yet to create an algorithm or a machine that can think in similar metaphors.

This is part of why there is so much hype around AI. When people talk about artificial intelligence, what they mean is a computer that really thinks. Forget the Turing test, because we have now passed it—but nobody believes that those chat bots are actually thinking. They are just implementing very specific models and making predictions about the right things: the right words to return in a text box or the right squeaks to make out of a speaker. But they are not self-aware; they don't have feelings. And they are not really thinking. We may build such systems one day, but we are nowhere near that right now.

One can, right now, go to Google or to Amazon and obtain voice recognition as a service—an algorithm that can recognize and parse spoken language—and one can put that to use for many different purposes. Once those companies have built services like that, there is really no reason to reinvent that particular wheel—they have much more voice data than anybody else. But as impressive as those services may be, that capability is very specific: it is trained on human voice, and finely attuned to human voice.

There is still no progress in the field to create a general-purpose AI algorithm or service that is capable of thinking in the flexible way that people can.

3. AI WINTER TO RENAISSANCE

"We weren't sure why, but we didn't understand how to build these systems after all."

This isn't the first time AI has run into a hype bubble. By the mid-1970s, there was widespread dismay at the state of AI research in general, and its failure to produce machine intelligence. The result of this disappointment was the first season referred to as an AI winter (coined for comparison with nuclear winter).

All of the early ambition and initial success had met with slow progress and crushing failure, and the idea that neural networks could usher in a new age of computing largely faded. Almost no graduate students were going into AI during that period—the prevailing thought was that it was a backwater, and nothing serious was going to come out of it. We weren't sure why, but we didn't understand how to build these systems.

So everyone went off and studied other parts of computer science. From databases to operating systems, many other areas were aggressively pursued. But no one invested in AI for a very long time.

Fast forward to today, and you can look around on any street and see someone whipping a phone out of their pocket and just talking to it. And they have pretty good confidence that the phone is going to do the right thing in response. It is simply breathtaking how good natural language interfaces are now.

And predictive systems: we now routinely look to computers to answer questions about both the present and the future. Is this credit card transaction fraudulent, or not? Will this patient need to be readmitted to the hospital after they're sent home, or not?

These systems are everywhere, and they are hugely impactful and powerful. That's a total shock. If you looked to the 1980s and early '90s, there was almost no work going on at all; no one believed these systems could be built. Fast forward just another 10 or 15 years, and all of a sudden they are everywhere.

So what on earth happened?

Today, we think nothing of carrying on a conversation with our phones or with virtual assistant devices in our homes.

3. AI WINTER TO RENAISSANCE (CONTINUED)

"Separating all of the hype from our real capabilities and what is critical for people who want to use this technology."

As powerful as it is, the AI we have today isn't really intelligence. It is pattern recognition by a class of algorithms that is excellent at pattern recognition. But each algorithm is still very special-purpose. For all the talk in the marketplace about cognitive computing, there is no real cognition at work inside of machines.

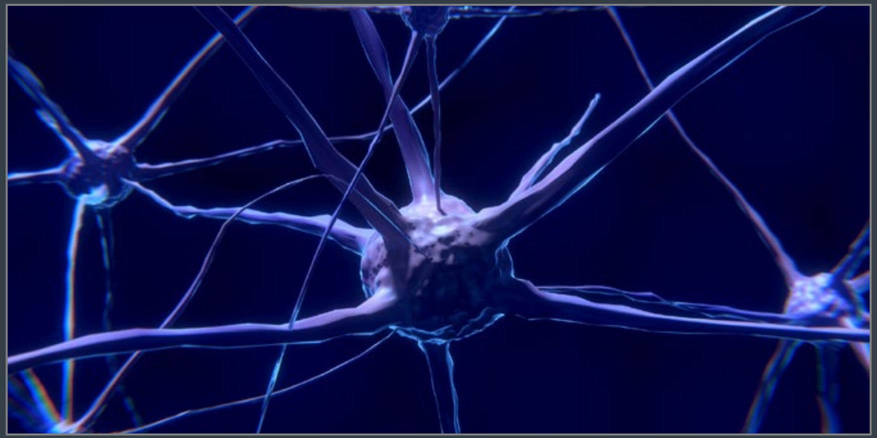
Early AI researchers wanted to recreate the structure of the human brain using computers. While we've made a lot of progress on the availability of large data sets and amazing levels of computing power, we haven't yet made enough progress on a functional understanding of the human brain. How does real consciousness emerge from two and a half pounds of gooey chemicals between our ears? We still don't know.

We have a much better model of the neuronal structure of the brain these days. The number of connections and variety of connections—and their change over time—is different than McCarthy and Minsky thought back in the day, and that evolving understanding has indeed informed some of the algorithmic study going on. But a lot of researchers in the field believe that, in order to build genuinely intelligent machines, we need to fully and fundamentally understand the structure and function of the human brain.

We need a fundamental new insight about ourselves in order to get to where computers are genuinely self-aware, and we haven't had that insight yet.

Separating all of the hype from our real capabilities and what is practical today is critical for people who want to use this technology. For this reason, some prefer to stick with the term machine learning over artificial intelligence. Others refer to AI and machine learning somewhat interchangeably.

Machine learning refers to a now classic set of algorithms that are pretty well understood from a technical point of view, which apply to problems we also understand today. That set of problems is growing as we get more experienced at deploying those algorithms, and as the computers get better.



We know more about the neuronal structure of the brain, but we still don't know how consciousness emerges.

4. YESTERDAY'S ALGORITHMS WITH TODAY'S POWER

"Today, we are essentially using techniques we thought of half a century ago, training them with data volumes and compute capacity acquired in the last decade, and solving the problems we originally intended to solve back then."

Until recently, we have not been able to use machine learning in fraud prediction at scale, for example, or to use natural language processing and sentence generation to converse with our devices. These applications are hugely transformative. And yet the algorithms that make them possible were, for the most part, first conceived in the early days of the field. It's the ability to deploy those algorithms that is all brand new.

So what changed?

The first thing is, back in the day, pioneers like Winograd, Minsky, and McCarthy didn't really have any data. If you're going to train a model to spot patterns, then big data is essential. The patterns you're looking for might be obscure, and you are going to need a lot of data to tease them out. Early AI researchers had to hand-code everything, and it was incredibly laborious to build data sets of just thousands of operations. Today, that's not a problem; we are generating data at a furious clip. We have an enormous amount of information available—and any time we want more, we can get it very easily, because seemingly everything has a sensor on it. And we can afford to store and use this data in ways we never could before.

The second thing is, we finally have enough computing power. Minsky was trying to build a 15-neuron model on an early iteration of ENIAC. Today, we can gang together thousands of vastly more powerful Dell or HP boxes with 16 CPUs each, and run really numerically complex algorithms exhaustively on our data. Thanks to Moore's Law, we have all the computing power we need.

Despite the AI winter and a season of disappointment, it turns out that the insights that founded the field were spot on, and a bunch of the early algorithms were actually pretty clever. We just needed to wait 50 years until there was enough computing power and data to put those insights into practice.

Today, we are essentially using techniques we thought of half a century ago, training them with data volumes and compute capacity acquired in the last decade, and solving the problems we originally intended to solve back then.

Even twenty years ago, it seemed impossible that we would have solid, general-purpose, speaker-independent voice recognition today. We had been working on that particular problem for the duration of entire professional careers without getting any better at it. And then all of a sudden—with lots of data and computing power, and these old machine learning algorithms—the problem is solved.

Who knows how quickly future advances against seemingly impossible problems will happen?

5. THE CURRENT STATE OF AI INNOVATION

"The biggest vendors... get to invest in building out new AI capabilities. Notice also that those companies have a disproportionate share of the data. That's why their algorithms and systems are better."

One of the intriguing things about AI and machine learning is that we're using many of the same techniques we've been using for decades, although thanks to improvements in compute and data availability they are becoming more useful. Common practices still include statistical and numerical fit algorithms such as random forests, Bayes, and so on. There has been innovation, but much of it has been in the detail, not at the high level.

One useful innovation is to use a dozen or so different models, let each of them make a prediction, and then use a consensus mechanism among them to draw out the best idea. When IBM Watson won Jeopardy, it used a very similar technique. The machine learning models that knew the answers (or rather, questions) in the Potent Potables category and the models that knew the Presidents category were totally different models. In a consensus approach like this, the answers can be evaluated and weighted in different ways for each question, based on the area of specialty and the level of consensus among the models, to yield the best answers.

The biggest vendors (like Facebook and Apple and Google)—with a disproportionate share of the money in the world—get to invest in building out new AI capabilities. Notice that those companies also have a disproportionate share of the data. That's why their algorithms and systems are better.

Companies of this ilk are indeed beginning to innovate by developing novel algorithms. For example, in 2014, Google acquired the London-based DeepMind research group that is pushing the boundaries of AI. But by and large, we are still at the beginning of this kind of high-level evolution.

6. AI AND YOUR BUSINESS

“If you are tempted by the promise of artificial intelligence, you can get started with practical problem-solving now...”

You are likely already seeing a huge transformation in your data centers and IT department even without AI. The wide availability of cloud computing and data storage means you have likely been building up massive-scale systems in your own data center. If you want to maximize value from those systems and take advantage of all that capacity to solve the right problems in the right order, then using machine learning algorithms to model your data and deliver valuable insights is naturally appealing.

If you are tempted by the promise of artificial intelligence, you can get started with practical problem-solving now, using available machine learning techniques. While truly cognitive computing is still a way off, tangible returns from deploying today’s models are a practical place to begin.

Without implying where to start first, here are a few examples of business problems that algorithms can help address:

- Maybe you want to get to know your customers better. You want to offer them services and products that are going to better suit their needs. Or you want to engage them better so that they continue to be your customers.
- Maybe you want to make predictions about what will happen next in your market. Today, most companies have furious streams of information landing in their data warehouses: the number of stock trades on the planet is skyrocketing; factory floor machines are all sensored up; online customer transactions and support conversations are leaving reams of text in their wake. You need to take the big data streaming in, and build connected systems that will give you reliable insight.
- Maybe there is a lot of risk in your business. Hackers are trying to break into your network. Or you’ve got employees who could be looking to steal secrets. We’re living in a GDPR world, so how can you be sure that you know where your personal data is?

Some business problems are not yet and will likely never be appropriate for machines to take over entirely. For example, we shouldn’t let algorithms have the final say on how to treat patients—but they may advise doctors on how to treat patients, while the doctors use their own judgment to make a final treatment decision. However, as computing power continues to grow, so does our ability to do more with modern algorithms; we are constantly expanding the use cases.

7. WHY YOU DO NEED EXPERT HELP

"People with the requisite skills in AI and machine learning are still relatively scarce, and commanding a premium in the market. Yet you need someone who understands the field to give you sound advice."

At Cloudera, we don't profess to have cognition inside our computers. What we do have is many hundreds of customers that are running mission-critical, predictive machine learning algorithms in production environments that provide them with solid advice for taking action that is driving business returns.

Even if you're not Google or Apple or Facebook, working on the cutting edge of developing new algorithms, you can still apply machine learning techniques to business problems that matter. The first step is to identify the data you have available—as well as any adjacent data you may be able to retrieve and pull in—and the business questions you can ask of that data, using currently available algorithms.

Ask yourself: "If I could learn about them, what patterns in my transaction flow, or in my call-center data, or in my employee hiring and decision-to-leave data, would make a tangible impact to my bottom line? What models would I like to build to help me answer good business questions? What other data could I go get? What have other companies done with similar data? What questions have they asked? What critical business problems matter to me as a decision-maker or as a purchaser that could be better informed by an algorithm?"

Once you determine which questions you want to ask, then there is a second set of questions to consider: "Which is the best algorithm, or suite of algorithms that connect? How should I train them, in a cost-effective way, to get the accuracy that I want?"

All of this requires some expertise. For one thing, over time, the data that you're analyzing in real time will drift away from the circumstances and data that your models were trained on. Periodically, you need to retrain your models in order to keep them accurate.

Furthermore, there are a host of nuanced ethical pitfalls in training and deploying machine learning. For example, when forecasting fraud in real estate loan applications, we must take care to avoid redlining or instilling bias in the algorithm via biased historical training data. For this reason, it's important to ask the right questions of available models and to train them with care.

People with the requisite skills in AI and machine learning are still relatively scarce, and commanding a premium in the market. Yet you need someone who understands the field to give you sound advice.

And you need a solid foundation in data management and analytics, as data science and machine learning are part of a connected workflow which lives across the data platform. You need the ability to efficiently ingest, store, transport, clean, prepare and catalogue data with context to be well positioned to build, deploy, and manage machine learning models efficiently, at scale.

8. THE FUTURE IS INDUSTRIALIZED AI

Enabling such a repeatable workflow means... abandoning one-off or ad-hoc approaches to project management and investment, and focusing on developing skills and best practices that translate from one team or business unit to another, or one company to another.

Successful implementation of AI is dependent on your organization's ability to get relevant data in the right place at the right time, process it, and develop useful models that can be deployed and managed in production. Until recently, this has happened mainly within a few pockets of innovation within the enterprise. But organizations are now looking to operationalize machine learning at scale, industrializing AI for use everywhere across the business.

This requires transformation at scale across several dimensions for the establishment of a repeatable, scalable workflow: from use-case identification to development to production deployment—and repeating the cycle all over again—learning continuously from previous outcomes, and optimizing for better efficiency. Enabling such a repeatable workflow means moving away from data and technology stack silos, abandoning one-off or ad-hoc approaches to project management and investment, and focusing on developing skills and best practices that translate from one team or business unit to another, or one company to another.

It all starts with defining your vision and strategy, then understanding how to prioritize investments. This requires a different way of thinking. For example, developing new and differentiated machine learning capabilities isn't like a deterministic engineering project where you know at the beginning about how long development will take, how much it will cost, and how much value it will drive. Machine learning projects are more like experiments. You can think of maintaining a portfolio of them, the way you maintain an investment portfolio, with some capabilities yielding more value than you expected and some yielding less. That can feel risky, but it's worth it when you remember that you can build on everything you accomplish to enable new possibilities—compounding value and continuing to build based on this approach.

Next, imagine a future state that looks more like what we might call an AI factory: where we'll have the organizational machinery for consistently and repeatedly discovering opportunities for AI and being able to act on them. Business and product leaders will be able to recognize new opportunities—they'll know what's possible as well as what's affordable or expensive, and be able to provide feasible ideas for data science teams to develop into solutions. An AI factory will enable the ability to rapidly deploy multiple use-cases while driving continuous learning and transformation across the business.

In this future, organizations are arranged around their abilities to recognize when machine learning automation would be useful in a business process, quickly experiment with different approaches, and easily deploy applications into production. But this type of enterprise needs a different kind of technical foundation—a platform for turning data into predictions, at any scale, anywhere.

8. THE FUTURE IS INDUSTRIALIZED AI (CONTINUED)

We're evolving our technology stack around our core data management DNA to design a next-generation, cloud-native platform for enterprise AI.

At Cloudera, we know that data management is the foundation for machine learning and data-driven automation. So we're evolving our technology stack around our core data management capabilities to design a next-generation, cloud-native platform for enterprise AI. Our platform will enable the complete machine learning life cycle for large teams, across massive data and heterogeneous compute, at enterprise scale.

And we'll continue to help our clients with a comprehensive approach to accelerating enterprise machine learning with expert guidance from Cloudera Fast Forward Labs. Because differentiated machine learning capabilities that drive the right business outcomes cannot be completely outsourced, we focus on helping clients with strategy and best practice development, knowledge transfer and upskilling, so clients are empowered to differentiate and own their future based on knowing their business and owning their data, skills and capabilities.

The opportunity is industrialized AI and the imminent challenge is scale. Let's futureproof and scale your business—together.

LEARN MORE