DefinedCrowd®

Part II

Toward Universally Benevolent AI

Developing machines that can reason like humans

By **James Whittaker**
Chief Strategy Officer

Abstract

The process of developing software has evolved and matured over the last human generation as the tech world has progressed through mainframes, PCs, the web and mobile platforms. However, one thing has remained consistent: code as the primary artifact of the software age. With the dawn of AI, there is a fundamental and unstoppable shift away from code toward data as the artifact that the rest of the software development process will revolve around.

This shift to a data-centric approach to application development allows AI to solve problems out of the reach of ordinary software. Unlike software, AI can reason in a way that is profoundly human. Technically, however, coding is still heavily involved and as a result AI remains the jurisdiction of technical people. But, there are major differences in the process of building AI and ensuring its safety and fairness that require much more participation by multidisciplinary stakeholders. Thus, the technical details are presented in such a way as to be digestible by anyone who wants to be part of this amazing future.

The Next Great Digital Age

The last few decades bore witness to a rapid transition from the paper-and-ink economy to the software economy. One by one the business processes that required humans to move paper were replaced with software -- first on mainframes, then PCs, then the web and finally the modern mobile apps we use today. Digital is now the new normal.

The progression of software through each of these form factors allowed us to solve increasingly sophisticated problems. For example, we could solve business problems like payroll or account management on a mainframe, but students couldn't do their homework on one. Nor could one deliver a PowerPoint presentation, write a resume or edit a photo. Those applications required PCs and the era of desktop apps was born.

But desktop apps were limited by the power of the machine they resided on. The popularization of the worldwide web allowed a new class of network apps to change the way people lived and worked. Apps for shopping, search and social networking appeared that broke the bounds of the desktop. Network software leveraged the power of both the local machine and remote servers; making apps possible that couldn't be built on standalone PCs, such as those for collaboration, dating and customer service.

By early this century, even the power of the network became too restrictive for software. Thanks to smartphones, software was able to break the physical barriers that confined it to offices, homes and internet cafes, and suddenly no one needed to wait for an internet connection to work or play on the road. Apps like mobile maps (no more printing driving directions!) and Pokémon Go would never have been possible without a portable form factor and an always-on data connection.

And now, right on schedule, the technology universe is shifting to yet another platform that will allow it to sink even deeper into the fabric of our lives. This time, however, it won't be built on any specific hardware. Instead, AI is being built on the vast quantities of digital data that humanity has amassed over the last 30 years. Nearly every aspect of human life from travel to weather to retail to medicine has been reduced to data and, thanks to the cloud and a collection of powerful pre-built algorithms (more on both of those technologies in a bit), the world's data is becoming a giant digital

The Next Great Digital Age

library for a new age of machines to learn what it means to live and work as humans on planet earth. The digitization of our lives has led to this moment where machines can process all that digital data to learn what we do, how we do it and begin to take over vast swaths of our work and leisure. Thanks to existing digital data, machines are in a better place to execute on humanity’s future vision than we are. Because machines are very good at data.

How Machines Learn

The transition from a code-centric world to a data-centric world requires a new set of processes and tools to build apps. In the code-centric world from mainframes to mobile, app behavior is programmed line-by-line and the finished product is a static binary capable only of obeying that programming. In the data-centric world, app behavior is learned by pre-existing algorithms that can react to changes in data and adapt new behaviors over time.

This way of learning is consistent with how humans learn. This means that the machines that will emerge over the next decade will be able to reason in very much the same manner we can reason, allowing them to solve problems out of the reach of software and compete with humans at certain types of tasks.

Consider, say, how we teach our toddlers geometric shapes and their names. We acquire some training material, maybe flashcards, and display those cards one at a time while carefully pronouncing the name of the shape as we do so: "triangle," "square" and so forth. Patient repetition takes care of the rest as the toddler forms a mental model of what it means to be a triangle or a square and how to associate those images with the spoken words that represent them.

AI learns in a remarkably similar manner. The flashcards would be stored and organized digitally. This is called source data or simply data. A data scientist would then play the role of teacher, carefully labeling each of the shapes with their corresponding name: triangle, square and so forth. Then an algorithm would process all that labeled data to create a mathematical model of shapes much like our mental model when we learn our shapes as toddlers. Which algorithm we choose (we'll get to this later) is a matter of experience. Some data is better suited to certain types of algorithms and the field of data science was created to advance scientific thinking on this newly important topic. Data, pre-existing algorithms and expert supervision are the primary tools of this new era.

Manually written code is no longer required to the heavy lifting, rather the focus has shifted to the collection, curation, labeling and modeling of data. This is what we mean when we say that the technology world is shifting from a code-focus to a data-focus. Data is the new code, the new paper-and-ink. Data is the medium that will define how humanity executes its future.

How Machines Learn

Consider: The right chess move is buried in the data that represents all prior chess moves and the universe of possible future moves. The number of people in a crowded room can be determined by data from sensors. The intent hidden among the words of a sentence can be extracted from the data of a million similar sentences all digitally encoded. For two generations of computer science, experts theorized that someday, with enough data to encapsulate human knowledge, paired with algorithms fast enough to accurately extract that knowledge, we'd enter a new golden age of AI. Turns out, those experts were right.

Humans have left a data trail about how they live, work and play. And machines are far better at finding the patterns in that data trail than we are.

All the history of the digital age has led to this point.

The Age of Artificial Intelligence

In philosophical terms, what is happening is that, thanks to all that data, algorithms can perform tasks that approximate human reasoning. Tasks that used to require a human (playing chess, counting the people in a room, understanding the meaning of a sentence), can now be performed by algorithms able to find patterns extracted from data about how humans perform the same task. If a task can be reduced to data, machines can perform that task.

And it is happening all around us:
Every time we type a query into a browser, we are using a data model: a vast store of websites, prior queries and known answers is tapped by Google's search algorithm to deliver results to our query. Every time we use mobile maps, geographic data along with the signals from all the cars and phones on that section of road are crunched by the routing algorithms.

Every time we type a sentence for a term paper, the data from untold term papers has created a model of good versus bad grammar to help us.

When we talk to Siri or Alexa, a model of human speech that has evolved from tens of millions of recorded utterances is consulted to determine what we said. Another model has learned to extract our intent from those words. And yet another model is consulted to respond with an answer.

The same is true for investment algorithms for banking and finance, shopping and recommendations engines for retail, scheduling algorithms for airport flight traffic, facial recognition for CCTV and building access control, and so forth. They all exploit the data of many thousands of people who performed those tasks while background software watched, carefully recording the data and built a digital model of that task. It is very hard now to live a life where some AI isn't either generating data from your activity or crunching data on your behalf.

It Starts With Data

Each of these examples of AI in everyday use is only the tip of the iceberg. AI is becoming part of most software projects and taking on oversized roles in places where software is falling short. AI's ability to churn through legal case precedence is making it very useful at assisting legal work and planning trial strategy. Data from millions of CT scans is being used by AI to find tumors and diagnose disease. AI is helping find new medicines and creating new vaccines. Data is even playing a role in creative fields by creating art and designing board games from data.

If a problem can be reduced to data, like search, maps or grammar checking, it is a candidate for AI. And when it comes to data, more is better.

Data is what makes Google's search engine better than any other search engine. A couple orders of magnitude more data is in Google's index and it receives far more daily queries to learn from.

It is also data that determines the scope of a product. Google Home is better at search due to the size of Google's information index. Alexa is better at shopping because of Amazon's depth of retail data about products and shopping habits. Each is better where they posses more data, and consequently have a huge head start in training algorithms to make sense of that data.

The role of the cloud

Despite being theorized as a concept since the 1950s, AI has only recently creeped into the consciousness of rank and file software developers. The reason, as you likely can surmise from reading this far, is because AI has been too data intensive for any prior stage of software development. AI requires so much data and so much processing power that older storage and compute technologies couldn't keep up. The algorithms that power AI are far too hungry for lone hard drives or even hard-wired server farms to satisfy their data and speed needs. For decades, AI has been quietly biding its time in anticipation of a technology to meet those needs.

Enter: the cloud.

It is no coincidence that the explosion of AI development postdates the arrival of cloud technology.

It Starts With Data

The cloud is the perfect place to gather data in large enough amounts to make it amenable to machine learning and, at the same time, provide the on-demand compute power that AI algorithms require. The cloud allows all the data associated with solving a problem to be co-located, indexed, hyper-organized and constantly kept up to date. Early purpose-built versions of the cloud enabled Google search algorithms to be the anticipatory magic they are today. Facebook's supernatural ability to connect users with content-needles in vast data-haystacks is also powered by their private cloud. It took public services like AWS and Azure to bring that capability to the rest of the world.

Indeed, it is the acquisition of data, through its collection, generation or purchase that starts every AI journey for tech teams. Without data, and a place to manipulate it, we would have been forever in the land of code-centric software.

Data primarily comes from two sources: devices (and the code that powers them) or people (specifically crowds of them doing repetitive things) performing actions and logging their information. Google got its web index from an algorithm crawling DNS servers making a list of URLs and it got its queries from humans typing into the search box. Amazon got its retail data by logging what users buy and where. Siri got its speech data from listening to what people are saying to it. There is no place more suited to this volume of data than the cloud.

Likewise, that people-counting app we talked about in part 1 of this manifesto needs data about floorplans and ingress, egress points. It needs rosters of authorized occupants and visitors. It needs data from motion and acoustic sensors as well as wireless access attempts by devices. The cloud is the perfect place to store and organize all those static and dynamic signals.

The algorithms that predict weather, the strength of hurricanes, the existence of exoplanets, the spread of pandemics, the presence of a tumor in a CT scan, and so forth, all require the vast amounts of historical data and sensor input that are best managed by AI and not software. If a solution requires insights based on data, then AI is the only non-human way to arrive at those insights.

Data allows AI to reason as if it were human and with enough of it, AI can solve problems that were once strictly the domain of humans. AI can assess traffic conditions and reroute cars far better than helicopters full of humans doing this by observation. It can count people in a building far faster and much more accurately than humans walking around with clipboards and mechanical counters.

It Starts With Data

For example, one of the earliest uses of AI, beginning in the 1980s, was to recognize human speech. Microphones were used to collect speech samples which were then digitized into usable data. AI algorithms poured over that data, recognizing the patterns that correspond to specific words. Then, human trainers would label those patterns, e.g., this pattern is the word "happy" and this other pattern is the word "days" and, voilà, the AI began to learn more words and could recognize full sentences.

Now imagine the amount of data one would need for that AI to learn every word in any given language. Imagine how much data it would need to learn every dialect, accent and context for all those words. It is no wonder applications like voice assistants had to wait for the cloud to become truly useful: that is a lot of data!

But AI is patient and with qualified linguists well-versed in technology continuing its training, speech AI is learning not only words, but their meaning as well. In fact, AI is so good at extracting meaning and understanding human intent that graphical user interfaces are being threatened. We are beginning to converse with machines in the same way we converse with other humans.

As simple as this process sounds, it took many decades to gather the data to get to the point where we talk to Siri and Alexa. The datasets that have led to these marvels have been painstakingly gathered from audio and video recordings and meticulously labeled and curated by data scientists. Indeed, the market for speech data is a harbinger of things to come. Data to train AI will become an industry all on its own with data owners, brokers, curators, maintainers and interpreters all employed in service of data. Data, it is said, is the new oil and the largest industry of the future will be the individuals and companies who refine data, of various forms and to serve various purposes, into knowledge.

From Data to Knowledge: Labels, Algorithms and Models

The process of developing AI has enough in common with the process of developing software that the two are often treated as synonymous. However, the latter is a well understood process of creating specifications and design documents, writing and testing code and then deploying the resulting app to the app store. Training AI is a much different process. Let's now look at this process in more detail.

Data collection and organization

Much of the power of AI lies in its ability to identify patterns in data. If AI can recognize objects in a camera shot, then it can pick out human faces and either count the people it sees (if that is its mission) or perhaps recognize a specific person and open a door for them. But this ability to recognize patterns and act on them requires a lot of data and practice sessions before mastery eventually occurs.

AI, of course, works on digital data: a specific real-world problem that has been painstakingly reduced to data and made available on a digital medium. AI researchers often talk about data acquisition (the collection or purchase of large data sets), data ingestion (the storage and representation of the data in a place and format that the AI algorithm understands) and a data quality step that allows data to be influenced, corrected and updated by a human data scientist.

For example, the recognition of human speech is a problem that linguists, data scientists and AI developers have been working on since the 1980s. Data is acquired by making recordings of people speaking. That data is ingested using a process (plain, ordinary software) to store new recordings and ensure they are properly formatted. Data quality is ensured by a human curator who can correct errors and otherwise ensure that the data is complete and representative of the humans that will eventually use the system.

There are two general ways that speech data is collected. Either through recordings of people in the wild having conversations or directed recordings from test subjects who are instructed what to say in an environment where ambient noise can be controlled. For the former, a human curator must listen to the recordings and attach labels that give the algorithm something to help it extract meaning. For example, the label "the speaker is asking for the latest presidential poll data" or "the subject of this conversation is football."

From Data to Knowledge: Labels, Algorithms and Models

Labeling data after the fact is very time consuming, as you might imagine, but it is also something machines can help with. Once the AI associates certain utterances with labels, it can help attach those labels on its own. In the latter case, data scientists direct test subjects what to say in advance so the labels are already in place. One can imagine how many samples Apple has of people saying "Hey Siri."

In both cases, a cloud-connected microphone performs the task of both digitizing the data and putting it in a place where AI can process it. Indeed, much of the data collected for processing by AI comes from sensors and devices capable of directly seeing the world as data, such as microphones or cameras that generally fall under the term "IoT," the internet of things. As such, it is almost impossible to separate the fields of IoT and AI as the former collects much of the data that AI processes.

Of course, humans play a role in this data generation as well, and the term crowdsourcing generally describes this aspect of data collection. Both IoT companies and crowdsourcing companies are generating vast amounts of data for AI and making it available to third parties via data platforms. Data is becoming a commodity that can be bought, sold and traded on public markets. Data is indeed proving itself to be the new oil.

Models and algorithms

Labels are simply informative tags that describe the meaning of specific datum. For example, we may add the tag "female face" to a photo that has a female face in it. Or we can add a textual label of the word being spoken in an audio sample. When we tell our toddler that a picture is a giraffe, we are conveying a label to the toddler that this is a giraffe. Labels help with the identification and classification of data to both humans and AI that are trying to learn a subject. Labels, in this sense, convey knowledge about a piece of data.

However, knowledge about a subject isn't the same thing as reasoning about a subject, which requires many layers of additional structure. A toddler may learn that a picture has a giraffe in it but still not be able to identify that giraffe as a land animal, a tall animal, a leaf eating animal and so forth. This kind of data inference and association requires more than labels. It requires structure and that's where data models come into play.

Models are more complicated as they relate to groups of data or indicate the commonalities in disparate datasets. Models describe how data relates to other data and encapsulates real world properties of the data.

From Data to Knowledge: Labels, Algorithms and Models

Knowing, for example, that lions, giraffes and wildebeests all have 'savannah' labels as their habitat helps both toddlers and AI to reason that they live in the same area and can be associated with Africa and even specific areas within Africa. Labels help with association and that leads to machine reasoning.

At their essence, models are data plus all the associations and insights that AI has been taught about that data. In other words, models are data that has been converted into knowledge. Humans often talk about "mental models" of how things work in the real world: we have a mental model about how airplanes can fly or how cash registers work without understanding every detail. AI models are similar except that AI can keep learning and tirelessly improving their models of the world. If you think voice assistants are good now, wait until their models of speech understanding begin to reach the levels of ours.

Models are constructed through a trial and error process with a great deal of participation from human data scientists and subject matter experts. After data is collected and labeled, data scientists choose the type of algorithm they think is best suited to extract the patterns from that data. Algorithms generally fall into a three categories: statistical models, stochastic models and layered models.

Statistical models are mathematical algorithms that have been around for ages and have been commonly used to forecast trends. They generally take data as batch input and try to classify it (the economy is more likely to expand than contract) or predict the next value in a time series (the hurricane is going to turn north). Upon the collection of more data, the entire batch is simply run through the algorithm again. Bayesian algorithms, regression algorithms and really the entire field of probability and statistics fall into this category. If one branch of math were to be crowned the precursor to AI, it would be statistics.

Stochastic models add structure and memory to the processing of data. Instead of outputting an answer or prediction, a model is constructed that is often depicted as a network or multi-dimensional matrix. Nodes in the network are connected when some relationship between the data that they represent is identified. The relationship can be causal (warmer water temps indicate a stronger hurricane), relative (all the common misspellings of a word) or anecdotal (higher volumes of speech often indicate anger). Stochastic models are more easily updated with new data and don't require batch processing as they remember what they have already learned. Clustering algorithms, Markov models and neural networks all fall into this category.

From Data to Knowledge: Labels, Algorithms and Models

Layered models are a type of stochastic model that separate and deepen the analysis of various parts of the data. For example, if an algorithm has identified facial features within the pixel data of a photo, a second layer of analysis would separate the pixels that represent, say, just the eyes within that face to determine its geometry. Another layer could focus on the nose and so forth.

Ultimately an eye model, nose model, mouth model, ear model, etc might be extracted and connected to the larger model of the face.

The eye model might allow us to determine the face is that of a primate. The nose model might allow us to say with some certainty it is a chimpanzee and the relationship between the two models might allow us to say what type of chimp. This type of analysis falls under the category of deep learning and gives AI researchers far more detailed data models to work with. It could even be used to find new species of chimps that don't fall into any known category. That's the value of such deep and detailed models.

AI Oversight

If the idea of a machine learning the same way a human learns doesn't scare you a little bit, then I have failed to land the full potential of this technology inside your brain. So, let me be clear: this stuff is scary. If the idea of someone hijacking your laptop camera and ransoming the recordings it makes of you causes you think twice about the dangers of software, then elevate that concern to an AI making those same recordings of you without requiring anything more than a picture of you online. Let that possibility sink in and you have the right level of concern about this technology.

AI is going to get really good at many of the tasks that humans are good at. This means it will threaten far more jobs than software, control more important decisions than software and rapidly expand software's society-wide influence. We really need to think this through. So, let's take a stab at doing just that.

There is bias in all our data

Human bias and discrimination is baked into the fabric of our society. Bias and discrimination, in the form of racism, sexism and intolerance is rampant across every culture, continent and country.
That bias exists in the data that AI will be processing. Data on what it takes to be a European business executive is going to be heavily skewed toward white males, not because they make better leaders, but because the data shows a historical prevalence of white males in positions of power and influence. AI trained to identify and recruit leadership candidates is going to be heavily skewed unless accommodations are made during data collection, labeling and modeling.

Likewise, an AI trained to predict crime and incarceration in the US is going to target black males because of inherent racism in the US criminal justice system. Getting AI to understand this is going to mean that model trainers must take care to extract as much of the bias from the data, or train our AI algorithms to take it into account.

Indeed, it is in this last idea where most of the promise of AI making the world a better place lies: technology that allows us to make AI see past race, gender, disability, sexual orientation and belief systems has the potential to level the entire playing field for society. As scary as AI might be on the surface, it actually has the potential to be truly unbiased and it's hard to imagine humans ever attaining that designation.

AI Oversight

AI can disenfranchise large classes of users

The cost of AI bias has the potential to leave large classes of humans behind. Technology that opens doors for some but not others, both literally and figuratively, is the default case because this world and the data that represents it, have generations of wrongdoing well documented. It is going to take a conscious effort by everyone involved in the training of AI to identify, remove or circumvent the darkness in our data.

Fast forward to 2030 and imagine what AI has managed to learn and master. Now ask yourself: is this a world that we can be proud of? The answer must be yes, and researchers and practitioners need to focus now on the steps needed to get there, so that a positive result for all humanity is achieved.

AI is really hard to audit

Compounding the problem of data bias is the problem of auditing AI's decision making. If you've ever said of another human being "what could they be thinking?" in questioning their decision making, then you understand the problem of auditability. The idea that someone could possibly think in a way contrary to your own, like in the polarized political climate in America, puzzles many of us.

Well, the neural networks that AI use to make decisions are often as complex (and will become more so) as our own gray matter networks and just as hard to audit. The people-counting AI we've been talking about is one such example: given three separate input streams of voice, motion and connectivity, how exactly does the AI come up with the result that there are 184 people in a room? Good question and just like those who oppose our political and social views … we may never know what they are thinking.

There is a lot of room in the AI world for new science and understanding. If we want to commit to a better future for humanity, its going to require a commitment to better AI.

DefinedCrowd®