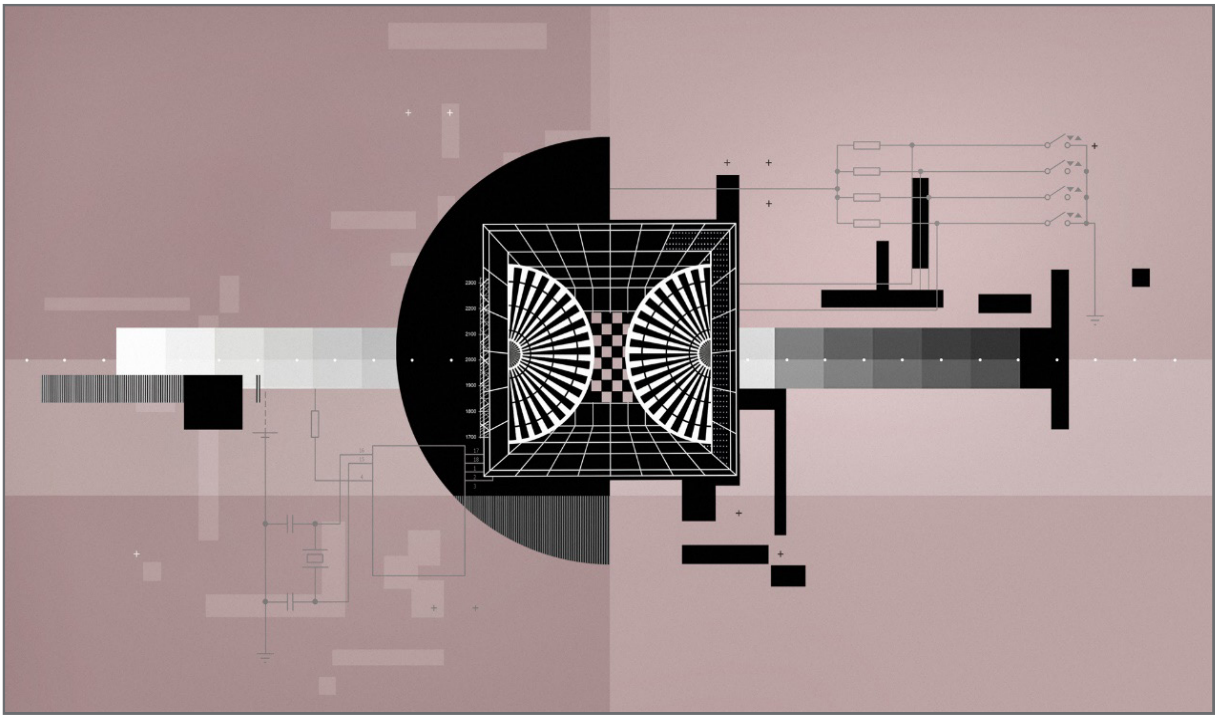
AI Does not Have to Be Too Complicated or Expensive for Your Business



Summary. For most companies that are interested in using AI, there isn't a clear model to follow. The approach to building AI used by massive internet companies like Amazon and Google just doesn't translate – most companies don't have overflowing troves of data they can...more
Despite the vast potential of artificial intelligence (AI), it hasn't caught hold in most industries. Sure, it has transformed consumer internet companies such as Google, Baidu, and Amazon – all massive and data-rich with hundreds of millions of users. But for projections that AI will create $13 trillion of value a year to come true, industries such as manufacturing, agriculture, and healthcare still need to find ways to make this technology work for them. Here's the problem: The playbook that these consumer internet companies use to build their AI systems – where a single one-size-fits-all AI system can serve massive numbers of users – won't work for these other industries. Instead, these legacy industries will need a large number of bespoke solutions that are adapted to their many diverse use cases. This doesn't mean that AI won't work for these industries, however. It just means they need to take a different approach. To bridge this gap and unleash AI's full potential, executives in all industries should adopt a new, data-centric approach to building AI. Specifically, they should aim to build AI systems with careful attention to ensuring that the data clearly conveys what they

need the AI to learn. This requires focusing on data that covers important cases and is consistently labeled, so that the AI can learn from this data what it is supposed to do. In other words, the key to creating these valuable AI systems is that we need teams that can program with data rather than program with code.

Why adopting AI outside of tech can be so hard

Why isn't AI widely used outside consumer internet companies? The top challenges facing AI adoption in other industries include:

1. Small datasets. In a consumer internet company with huge numbers of users, engineers have millions of data points that their AI can learn from. But in other industries, the dataset sizes are much smaller. For example, can you build an AI system that learns to detect a defective automotive component after seeing only 50 examples? Or to detect a rare disease after learning from just 100 diagnoses? Techniques built for 50 million data points don't work when you have only 50 data points.

2. Cost of customization. Consumer internet companies employ dozens or hundreds of skilled engineers to build and maintain monolithic AI systems that create tremendous value – say, an online ad system that generates more than $1 billion in revenue per year. But in other industries, there are numerous $1-5 million projects, each of which needs a custom AI system. For example, each factory manufacturing a different type of product might require a custom inspection system, and every hospital, with its own way of coding health records, might need its own AI to process its patient data. The aggregate value of these hundreds of thousands of these projects is massive; but the economics of an individual project might not support hiring a large, dedicated AI team to build and maintain it. This problem is exacerbated by the ongoing shortage of AI talent, which further drives up these costs.

3. Gap between proof of concept and production. Even when an AI system works in the lab, a massive amount of engineering is needed to deploy it in production. It is not unusual for teams to celebrate a successful proof of concept, only to realize that they still have another 12-24 months of work before the system can be deployed and maintained.

For AI to realize its full potential, we need a systematic approach to solving these problems across all industries. The data-centric approach to AI, supported by tools designed for building, deploying, and maintaining AI applications – called machine learning operations (MLOps) platforms – will make this possible. Companies that adopt this approach faster will have a leg up relative to competitors.

Data-centric AI development

AI systems are made up of software – the computer program that includes an AI model – and data, the information used to train the model. For example, to build an AI system for automated inspection in manufacturing, an AI engineer might create software that implements a deep learning algorithm, that is then shown a dataset comprising pictures of good and defective parts, so it can learn to distinguish between them.

Over the last decade, a lot of AI research was driven by software-centric development (also called model-centric development) in which the data is fixed, and teams attempt to optimize or invent new programs to learn well from the available data. Many tech companies had large datasets from millions of consumers, and they used it to drive a lot of innovation in AI.

But at AI's current level of sophistication, the bottleneck for many applications is getting the right data to feed to the software. We've heard about the benefits of big data, but we now know that for many applications, it is more fruitful to focus on making sure we have good data – data that clearly illustrates the concepts we need the AI to learn. This means, for example, the data should be reasonably comprehensive in its coverage of important cases and labeled consistently.

Data is food for AI, and modern AI systems need not only calories, but also high-quality nutrition.

Shifting your focus from software to data offers an important advantage: it relies on the people you already have on staff. In a time of great AI talent shortage, a data-centric approach allows many subject matter experts who have vast knowledge of their respective industries to contribute to the AI system development. For example, most factories have workers that are highly skilled at defining and identifying what counts as a defect (is a 0.2mm scratch a defect? or is it so small that it doesn't matter?). If we expect each factory to ask its workers to invent new AI software as a way to get that factory the bespoke solution it needs, progress will be slow. But we instead build and provide tools to empower these domain experts to engineer the data – by allowing them to express their knowledge about manufacturing through providing data to the AI – their odds of success will be much higher.

Make building and using AI systematic and repeatable

The shift toward data-centric AI development is being enabled by the emerging field of MLOps, which provides tools that make building, deploying, and maintaining AI systems easier than ever before. Tools that are geared to help produce high-quality datasets, in particular, hold the key to addressing the challenges of small datasets, high cost of customization, and the long road to getting an AI project into production outlined above. How, exactly? First, ensuring high-quality data means that AI systems will be able to learn from the smaller datasets available in most industries. Second, by making it possible for a business' domain experts, rather than AI experts, to engineer the data, the ability to use AI will become more accessible to all industries. And third, MLOps platforms provide much of the scaffolding software needed to take an AI system to production, so teams no longer have to develop this software. This allows teams to deploy AI systems – and bridge the gap between proof of concept and production weeks or months rather than years.

The vast majority of valuable AI projects have yet to be imagined. And even for projects that teams are already working on, the gap that leads to deployment in production remains to be bridged – indeed, Accenture estimates that 80% to 85% of companies' AI projects are in the proof-of-concept stage. Here're some things companies can do right now:

1. Instead of merely focusing on the quantity of data you collect, also consider the quality, make sure it clearly illustrates the concepts we need the AI to learn.
2. Make sure your team considers taking a data-centric approach rather than a software-centric approach. Many AI engineers, including many with strong academic or research backgrounds, were trained to take a software-centric approach; urge them to adopt data-centric techniques as well.
3. For any AI project that you intend to take to production, be sure to plan the deployment process and provide MLOps tools to support it. For example, even while building a proof of concept system, urge the teams to begin developing a longer-term plan for data management, deployment, and AI system monitoring and maintenance.

It's possible for AI to become a thriving asset outside of data-rich consumer internet businesses, but has yet to hit its stride in other industries. But because of this, the greatest untapped opportunity for AI may lie in taking it to these other industries. Just as electricity has transformed every industry, AI is on a path to do so too. But the next few steps on that path will require a shift in our playbook for how we build and deploy AI systems. Specifically, a new data-centric mindset, coupled with MLOps tools that allow industry domain experts to participate in the creation, deployment and maintenance of AI systems, will ensure that all industries can reap the rewards that AI can offer.