5 Things to Keep in Mind When Using Data for Artificial Intelligence

Data is one of the most important strategic assets for companies in the emerging data-driven and AI-powered economy. Data is needed to measure the efficiency of business strategies and draw insights from its operations but also to train machine learning algorithms. Getting data is not a problem for companies, the question is can they get the right kind of data and can that provide them with a much desired competitive advantage.

Related: Want to Be More Like Amazon? Start By Making Your Startup More Data-Driven.

Many companies do not realize that they are sitting on a pile of bad or dirty data. This data contains a lot of missing fields, has wrong formatting, numerous duplicates, or is simply irrelevant information. IBM research estimated that the annual cost of bad data for the U.S. economy is a whopping $3.6 trillion. Still, many managers have certainty that they are sitting on a goldmine of data when in reality they have nothing valuable.

I interviewed Sergey Zelvenskiy, who is an experienced machine learning engineer over at ServiceChannel, where he automates facilities management processes using artificial intelligence. We talked about common misconceptions when it comes to the good/bad data dichotomy and what companies should be focusing on when building AI products.

As Zelvenskiy says, “The data that companies have may not necessarily be bad, it is just likely incomplete to solve the problem. There is a chicken and egg problem here. The original system is usually built to collect the data needed for human-

driven solutions and moving it to an AI driven solution might require filling of the gaps. While a human can quickly assess these and fix the problem, the automated system needs automated ways to wrangle the data."

Focus on the product.

Finding good data should start with a product itself. To get good data, companies should design products that provide the right incentive for the users to contribute their data. Good usability and user experience will encourage users to contribute valuable information.

Related: Artificial Intelligence Is Likely to Make a Career in Finance, Medicine or Law a Lot Less Lucrative

You can always strive for the user-in-the-loop model, in which users have to give away their data in order to use the features of your product. This is precisely how Google and Facebook get tons of data in exchange for their services. Users are not even aware that they are giving away their data absolutely for free to power advanced machine learning algorithms and continually improve the software.

The best way to build a great product is by delivering iterative improvements while gathering the much-needed data. As Zelvenskiy says, "You can see this with the evolution of Amazon Alexa. The team behind it realized the difference between general speech recognition and the ability to recognize a simple set of predefined commands. While many other companies struggled with the adoption of general speech recognition and the capability to maintain the conversation, Alexa team focused on a simple set of commands and simple scripted dialogues."

The Alexa team did it right by shipping a very simple solution at a low price and conquered the market. Focusing on the specific simple use case and perfecting it wins the end game.

Related: 5 Reasons Machine Learning Is the Future of Marketing

Target the right types of data.

Let’s take the company that wants to build a robot that will automatically put library books on the shelves. It has plenty of data about the actual book content, it knows the names of the authors and the year the book was published. But, in reality, this data is not sufficient for an automated arrangement of the books.

The robot can use the existing data only to find the proper shelf for the book. But, it doesn’t know the measurements of the book, so it’s hard for the robot to figure out if the book will fit on the shelf.

The company never thought of collecting this information because the library staff could easily figure out if the book fits the space. Now this company needs a completely new data set, which it doesn’t have. This means the company has to equip a robot with some way of assessing the book measurements instead. While this is not impossible, the project budget and timeline will change.

That’s why you should always ask yourself if you have the right type of data that is helpful to solve the problem.

Related: Top 10 Best Chatbot Platform Tools to Build Chatbots for Your Business

Understand the limitations.

Often, companies feel that all machine learning engineers have the same magical wand, that solves all data-related challenges. That cannot be further from the truth. Going back to the library example, the ability to automatically assess the size and weight of physical objects would require a very different set of skills and capabilities. People or

systems who can train the robot to find the right shelf, are different from the people or systems capable of building the abilities to measure and weight the books.

This kind of resource planning should start at the beginning of the project and not when the robot is destroyed under the pile of books that did not fit the shelf.

Related: A Humanoid Robot Called Sophia Mocked Elon Musk After Being Asked About the Dangers of AI

Utilize existing expertise.

Artificial intelligence can only do it better after the hard work by the team of engineers and subject matter experts is done. The development of an intelligent solution needs expert inputs to understand and help interpret the existing data and to figure out the principles they use to solve the problem.

Even the latest breakthrough of DeepMind's AlphaGo Zero is not an absolute showcase that we don't need human experts entirely. The rules of the Go game are well-defined and cannot be broken by the opponent. Even though the machine was not trained by human experts, the rules of the game were programmed into the code, so it can play against itself to build up the skills. The engineer who built the software became an expert in the rules of the game before programming it.

According to Zelvenskiy, "In the case with AlphaGo Zero, we don't have a dedicated expert because the playing field is so well-defined that one can learn the complete set of rules in one evening. In real life, an engineer can hardly spend an evening and become an expert in the supply chain, privacy laws or turbine engineering. In general, an AI project either needs a well-defined set of unbreakable rules or a labeled data set. Usually, there is a little bit of each and figuring out how to combine the pieces of this jigsaw puzzle still requires expert

input.”

Zelvenskiy added, “Don’t get me wrong, there are success stories when a team of engineers successfully solves the puzzle by obtaining the right data set and learning just enough rules of the game. Yet, we depend on survivorship bias here.”

Related: 10 Artificial Intelligence Trends to Watch in 2018

Manage data and close the loop.

One day your application might start to generate large volumes of data as it gets more popular. To avoid running into a data mess, you should introduce efficient data warehousing strategies from the very beginning. No matter what data platform your company chooses, you should put in place the efficient process of data collection, cleansing and data wrangling at each stage of data acquisition process.

Once you have a good product, a constant inflow of data and an efficient data management infrastructure, it will be easier to create a self-fulfilling prophecy of good data.

Leveraging the data provided by your product’s users can improve AI platforms and application features and encourage customers to contribute even more good data. This will create a self-sustaining system of data generation that will turn your company a truly data-driven enterprise.

https://www.entrepreneur.com/article/307019