What is the Relationship between Data Science and Artificial Intelligence? – My View

Bhavani Thuraisingham

February 15, 2020

Over the past decade the world has heard a lot about Data Science and especially Artificial Intelligence (AI). Whatever we read and watch, AI is mentioned in most pages or scenes. There have been many debates/articles about the relationship between Data Science and AI and they create even more confusion. This past Christmas while in vacation in Bristol England, where I started my married life and graduate education almost 45 years ago, I had the time and opportunity to read and reflect on how Data Science and AI have evolved over the past several decades and explore the relationship between the two. Having worked in both areas, especially as they relate to cyber security and some other applications such as process control systems and networks for almost 35 years, I believe that I have arrived at a reasonable conclusion on what the two fields are about and their relationship to each other. So, here is my view.

Well before we had Data Science, we had Database Systems and AI since the 1950s. While the origins of AI go back to the seminal paper by Alan Turing in 1951 "Can Machines Think?", early database systems in the late 1950s were based on the network model (e.g., the work of the Codasyl Group). Both fields developed independently with expert systems and rule-based systems dominating AI in the 1960s and 1970s and relational database systems dominating database systems in the 1970s (while IBM's IMS database system developed in the 1960s was quite popular in the 1970s). I remember attending AAAI 1988 in St Paul, MN and much of the discussion was on expert systems. Furthermore, Machine Learning (ML) was developing as a sub-area of AI at that time. In the meantime, efforts to integrate AI and Databases began in the late 1970s with Logic databases and Intelligent database systems where reasoning and inferencing were built into database systems. This integrated intelligent database management was a major area throughout the1980s and into the early 1990s. While database systems research continued to explode throughout the1980s and 1990s, AI was not progressing as much and did not meet expectations. For example, while MCC's Cyc project had high hopes, the objectives of the project were not met.

As the amount of data to be managed grew larger and larger especially with the WWW, the word data management was coined around the late 1980s and early 1990s. It included not only database management but also managing the large amounts of data that included cleaning the data and ensuing the integrity of the data. Soon in the early 1990s, the goal was to extract nuggets from the data so that intelligent decisions could be made. Statistical reasoning techniques (borrowed from Statistics) and Machine learning techniques (borrowed from ML) were being applied to the data to extract useful patterns as well as predict new patterns. This field came to be known as Data Mining. Data Mining grew leaps and bounds between 1995 and 2005 and provided corporations the ability to analyze the data and extract useful information. At the same time data warehouses that clean and aggregate the data to answer statistical queries became very popular. Data warehousing was considered by many to be the first step towards data mining. That is, one has to get the data ready to mine it and extract useful results.

I remember attending the first KDD (Knowledge Discovery in Databases) conference in Montreal in 1995 and the debate at that event was who owns data mining? Is it database management, statistics or machine learning? Each group claimed that they owned it. I believe it was then decided that data inning belonged to all three areas as you need not only the correct data but also statistical and machine learning techniques to reason, learn and extract useful nuggets. However, in the early 2000s, DARPA's TIA (Terrorism Information Awareness) project gave a lot of negative press for Data Mining and so the United States Congress put a moratorium on Data Mining. Subsequently the name Data Analytics was used instead of data mining, Nevertheless the data mining area continued to thrive and flourish among the researchers.

In the 2000s the AI researchers were trying to advance machine learning by introducing many layers of learning in neural networks. This approach worked. Therefore, neural network research that was more or less dormant for some time soared with the concept of deep learning. By 2011 IBM released Watson and the popularity of ML was worldwide. Around that time, it was also possible to use cloud and high-performance computing to manage and analyze data. Data management, data warehousing and data mining/analytics all merged into a field called Data Science and was closely intertwined with the term Big Data. Data Science included both getting the data ready for analytics as well as the analyzing the large amounts of data which included both statistical reasoning and machine learning. Essentially one could consider Machine Learning to be a part of Data Science. Data Science is still very popular as without having the proper data you cannot apply the learning techniques effectively and produce accurate results.

While students came in plane loads to learn Data Science and Machine Learning especially starting in the 2010s, Artificial Intelligence which was by then over 60 years old became hugely popular partly because the AI researchers always felt that ML was part of AI and ML had its origins in AI. After all, it was the AI researchers who invented deep learning in the 2000s. A major area of AI has always been ML. But AI also includes some less popular areas today such as expert systems and planning systems which were major disciplines of AI in the 1960s, 1970s and 1980s. AI also includes other areas such as multiagent systems and robotics. But it is the ML part of AI especially with deep learning that really took off in the 2010s.

So back to the question "What is the relationships between Data Science and AI?" I believe that Data Science and AI both intersect at Machine Learning. Data Science also has other areas such as data management and data warehousing while AI has other areas such as planning and robotics. So now we have the Machine Learning debate. Who owns Machine Learning? The AI people say its them because ML has always been part of AI while Data Science people say its them as ML was part of Data Mining. But I believe it belongs to both areas. However, since the popularity of AI now has somewhat exceeded that of Data Science, the common belief is that Machine learning is part of AI. After all it is Machine learning that has given so much clout to AI and it is the AI researchers who developed deep learning. However, one cannot forget that without cleaning, organizing and managing the data efficiently, one cannot have ML and Modern AI. So, once again here is my view: MACHINE LEARNING IS AT THE INTERSECTION OF DATA SCIENCE AND ARTIFCIAL INTELLIGNCE AND BELONGS TO BOTH FIELDS.