Artificial Intelligence and Big Data: Good for Innovation?

Renu Kumari\textsuperscript{1}, Deepa Sonal\textsuperscript{2}

\textsuperscript{1}Asst. Professor, \textsuperscript{2}Lecturer, Department of Computer Science, Patna Women's College, Patna University, India

Abstract— Artificial intelligence is firmly embedded throughout the economy. Financial services firms use it to provide investment advice to customers, automakers are using it in vehicle autopilot systems, technology companies are using it to create virtual assistants like Alexa and Siri, and retailers are using artificial intelligence (AI) together with customer’s prior sales histories to predict potential purchases in the future.

Big data is being generated all the times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it.

The most dramatic advances in AI are coming from a data-intensive technique known as machine learning. Machine learning requires lots of data to create, test and “train” the AI.

Thus, AI is becoming more important to the economy, so data is important too.

Keywords: Artificial intelligence, Big Data, digital data, volume, velocity, variety

I. INTRODUCTION

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information. Big data is often characterized by three Vs: the extreme volume of data, the wide variety of data types and the velocity at which the data must be processed.

Although big data doesn’t equate to any specific volume of data, the term is often used to describe terabytes, petabytes and even exabytes (1 million terabytes) of data captured over time.

Big Data can offer great insights with the help of Artificial Intelligence (AI). Artificial Intelligence deals with the study and development of software and machines that can imitate human-like intelligence and it is a branch of computer science that is extremely technical. Artificial intelligence’s most common application is about finding patterns in enormous quantities of data. Smaller more homogeneous fixed data sets will not serve the purpose as pattern may not be evident in them. This allows companies to automate and improve complex, descriptive analytical tasks, which would be extremely labor intensive and time consuming if carried out by human beings.

The aim of this paper is to explore the opportunities of Big Data focusing on applications of Artificial Intelligence to Big Data problems. This paper begins with a brief overview of Big Data followed by a section discussing the application of Artificial Intelligence to Big Data and limitations and issues of Big Data and Artificial Intelligence.

II. BIG DATA: AN OVERVIEW

The use of term “Big Data” can be traced back to the discussion of handling huge groups of data sets in both academia and industry during the 1980’s. Michael Cox and David Ellsworth were the first to use the term Big Data literally, referring to using larger volumes of scientific data for visualization (the term large data has also been used) [Cox and Ellsworth, 1997].

The first formal academic definition appears in a paper submitted in July 2000 by Francis Diebold of University of Pennsylvania in his work of econometrics and statistics (2000):

Big Data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology. In this new and exciting world, sample sizes are no longer fruitfully measured in “number of observations”, but rather in, say, megabytes. Even data occurring at the rate of several gigabytes per day are not uncommon.

Gartner defines Big Data as follows:

Big Data is high-volume, high-velocity information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. [“Gartner IT Glossary,” n. d.: Lapkin 2012]

Tech-America Foundation defines big data as follows:

Big data is a term that describes large volume of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information. [Tech-America Foundation’s Federal Big Data Commission, 2012]
III. BIG DATA CHARACTERISTICS

The big data characteristics can be elaborated as follows:

1. **Volume:** This characteristic defines the quantity or magnitude or amount of data. The magnitude of available data has been rising at an increasing rate. This applies to both companies and to individuals. A text file is a few kilobytes, a sound file is a few megabytes while a full length movie is a few gigabytes. New data sources are added on nonstop basis. Facebook’s databases ingest approximately 500 terabytes of data each day [Smith, 2014].

2. **Variety:** Variety refers to the wide range of formats in which the data are being generated and are stored. Big Data extends beyond structured data to include semi-structured data and unstructured data of all varieties. Different applications, processes and systems generate and store data in varied formats. Such as data generated from sensors, devices, whether satellites, set-top boxes, surveillance/CCTV cameras, traffic cameras etc. These different formats of data cannot be stored in structured relational database systems.

3. **Veracity:** Veracity relates to the accuracy, reliability/trustworthiness, applicability, noise, bias, abnormality, completeness, and other quality aspects of big data.

4. **Velocity:** Velocity refers to the speed/rate at which the data are created/generated, transmitted, stored, processed, analyzed, accessed and visualized.

IV. APPLYING ARTIFICIAL INTELLIGENCE (AI) TO BIG DATA

Unleashing AI on big data can have a noteworthy influence on the role data plays in conducting business, analytical and decision making. Big Data is not component of Artificial Intelligence [Umbler Corp., 2015]. However, the two are entwined: AI provides the large scale analytics necessary to extract meaning and value from big data, while the big data provides the knowledge required for AI to continue to learn and evolve-or to become more intelligent [Umbler Corp., 2015]. In other words, AI offers the technology and methodology for better understanding of the ever growing amount of data.

Consumer Internet companies are in hurry to build out their AI talent and acquire the most advanced machine learning system. Some of the major acquisitions from AI field that occurred in recent month [Cooper, 2014]:

- Facebook launched a new research lab to dedicated entirely to advancing the field of AI so that they can predict what consumer will do in the future.
- Google acquired DeepMind, a company that build learning algorithm for e-commerce, simulation and games, for $US400 million.
• LinkedIn acquired Bright, a company that focused on data-and algorithm-driven job matches, for $US120 million—its largest acquisition to date.
• Pinterest acquired Visual Graph, a company that specialized in image recognition and visual search. Visual Graph helped build Google’s first machine vision application to improve image search.

Artificial Intelligence and Big Data can be used for fraud management. Neural networks are analytics that learn to distinguish complex patterns of behavior (in customer transactions, network activity, etc.). Neural network have been widely deployed in fraud management because they excel at swiftly spotting abnormal data patterns within large amount of transaction data.

Social networking experiences are becoming increasingly centered around photos and videos [Cooper, 2014]:
• Facebook users upload 350 million photos each day [Cooper, 2014].
• Snapchat users share 400 million “snaps” each day. [Cooper, 2014].
• Instagram users upload 55 million photos each day. [Cooper, 2014].
• 30,000 are uploaded on Flicker every minute. [Holland, 2015].

V. LIMITATIONS AND ISSUES OF BIG DATA AND ARTIFICIAL INTELLIGENCE

Irrespective of how powerful and how much value big data brings, it has its limitation. Big data has its distinctive characteristics that can provide decision makers with more timely, rich information, but without the specific context, data on its own can never tell the whole story. Like any data, big data is not a panacea to solve all questions for all organization.

With big data there will also be dirty data, with potential errors, incompleteness, or differential precision. AI can be used to identify and clean dirty data or use dirty data as a means of establishing context knowledge for the data. For example, “consistent” dirty data might indicate a different context than the one assumed. In short GIGO is applied when AI is applied on dirty data.

Current AI algorithm sets are often non-standard and primarily research based. Algorithm might lack documentation, support and clear examples.

VI. CONCLUSION

Artificial intelligence and Big Data mesh well. Artificial Intelligence and Big Data when put together opens up innumerable opportunities for resolving the problems faced in the modern society in 21st century and beyond. Gains from Big Data and AI will be governed by right analytic practices, clear understanding of business requirements that is – what AI is aiming to achieve from the data, relevant data and active involvement of and collaboration amongst stake-holders-data scientists, domain/subject matter expert, business and data analysis and top management support for endeavor.

REFERENCES