On the Implications of Artificial Intelligence and its Responsible Growth

Harsha Devaraj[1], **Simran Makhija**[2], **Suryoday Basak**[3]
[1]University of Michigan Ann Arbor
[2]Astroinformatics Research Group, IEEE Computer Society Bangalore Chapter and Center for Mathematical Modeling and Simulation (CAMMS).
[3]Department of Computer Science and Engineering, The University of Texas at Arlington

ABSTRACT
As a set of technologies, Artificial Intelligence (AI) has received growing interest from a variety of fields. However, many fundamental questions about AI are still mysteries to the everyday person. This paper seeks to address the history of AI, the current state of the field, important distinctions between related fields, misconceptions birthed by popular media, and irresponsible applications of AI. The authors believe this basic understanding of AI and its shortcomings is vital, and the advancement of the field should be matched by the advancement in public understanding of AI.
Keywords: Artificial intelligence, Machine learning, Ethics.

Correspondence

Suryoday Basak
Department of Computer Science and Engineering, University of Texas at Arlington, Box 19015, Arlington, TX, USA.
Email: suryodaybasak@gmail.com

Received: 15-02-2019
Revised: 11-06-2019
Accepted: 16-09-2019
DOI: 10.5530/jscires.8.2.21

INTRODUCTION

Almost sixty years after its inception, applications of AI are growing like never before. Corporations are trying to automate large-scale tasks using AI, social media content is now being filtered based on users' interests, smartphones are always suggesting something new for users to try, Google Maps is using analytics to optimize routes; the list is almost endless. Though AI has become an omnipresent helping-hand that assists us in many ways, the negative connotations and misconceptions that are associated with it have grown as well. Many of this stems from inaccurate representations in pop-culture and scifi movies; some of it comes from real examples of the harmful effects that AI applications have had in the past. Rational and irrational fears muddle together and many people are also unaware of exactly what aspects of their lives are already being influenced by AI.

This paper seeks to contextualize AI in history, and to do that, we begin at the birth of the field.

ARTIFICIAL INTELLIGENCE: A BRIEF HISTORY

Artificial Intelligence emerged as a field of research after a workshop in Dartmouth College in 1956.[1] The goal was to automate tasks in the way that humans would solve them. The attendees included Marvin Minsky, Herbert Simon, John McCarthy, Allen Newell, and others: some of the most influential minds in the conception of the field. At the time, they predicted that near-human intelligence could be achieved in about twenty years, and could be simulated in machines. However, they soon realized that the full scope of human intelligence was too vast to be simulated; till today, it remains the ultimate goal of AI. Nonetheless, their efforts were groundbreaking from both a philosophical, as well as technological viewpoint.

After this initial boom in interest, the second half of the 1970s (referred to as the AI winter) saw developments in AI widely reduced due to cuts in funding. Interest was soon revived due to the fusion of AI techniques into more critical business systems, more popularly known as expert systems[2]. When the potential of machine learning and pattern recognition[3] was explored with the goal of automating decision-making systems in business, interest in AI was revitalized.

In the 21st century, the progress of AI was fueled by various factors:

1. Rapid advances in computer hardware
2. Accumulation of large volumes of data
3. Spread of information due to the Internet
4. A pressing need to automate mundane tasks so that we, as humans, can focus more on conceptual and fundamental ideas
5. For the progress of businesses by exploiting preferences of consumers from historic data

6. Ease of communication with machines
7. Need to explore in fields where human-eye fails to observe patterns

The techniques used in AI, however, were not born with the workshop in Dartmouth: it is interesting to note that some of the fundamental ideas, which include least-squares regression, convex optimization, interpolation, etc. were in use many years prior. Regression had been widely used in physics and economics before they were popularized in machine learning (in fact, astronomical research benefited from regression analysis in the 19th century!) Convex optimization, which is at the heart of neural networks (perhaps the most popular AI idea in use) has widely been used to optimize production based on constraints on resources, etc. Thus, we see that the mathematical models that serve as the building blocks of AI were in use much before AI was established.

CURRENT STATE OF AI TECHNOLOGY AND RESOURCES

At present, AI is experiencing tremendous growth for the following reasons:

1. AI has found applications in speech recognition, searching patterns, interest estimation etc. These are widely being used by Facebook, Twitter, Google, Apple, and others in a variety of applications.
2. Fields such as bioinformatics, astroinformatics, geoinformatics, finance tech, etc., which accumulate data at a rapid pace, have developed machine learning methods to parse this data for valuable knowledge. These fields contribute to the growth of AI, which in turn helps the advancement of the field.
3. Big-data frameworks have been developed, such as Hadoop, Apache Spark, etc., which facilitate the process of analyzing large volumes of data. Open source libraries are widely available for languages like Python, R, MATLAB, C++, Java, and are being used by developers and researchers from various fields.
4. Educational resources on AI are widely available: various books, online courses, video resources, papers, etc. are introducing more people to field every day.

The factors mentioned above do not work in isolation from each other. Knowledge materializes into new methods and applications, which attract new practitioners, who build upon existing knowledge: and the cycle continues.

The nature of tasks being solved by AI indicate the kinds of techniques that should be used, and broadly, the different types of methods that exist within AI fall into the following categories:

1. **Supervised Learning:** Supervised learning involves learning a mapping between an input and an appropriate output. The training data in supervised learning are input-label pairs, (x, y), where x is a set of features and y is the appropriate mapping (or *target value)* for x. Based on the type of values that y can have, supervised learning methods are further categorised into two types:
   (a) *Classification:* When target values are categorical, or are class labels, the supervised learning task is termed classification.
   (b) *Regression:* When target values are continuous, the supervised learning task is termed regression.

1. **Unsupervised Learning:** In contrast to supervised learning, in unsupervised learning, the training data is not found with a target value. Instead, the tasks that are associated with unsupervised learning usually involve discovering patterns that are inherent to the inputs. A large family of unsupervised learning methods is *clustering,* wherein the task of the machine learning algorithm is to discover clusters in data that are adequeately different from each other.
2. **Semi-supervised Learning:** Semi-supervised learning techniques are usually a hybrid of supervised and unsupervised learning, as the name suggests, and makes use of unlabelled data in conjunction with labelled data to bolster the accuracy of the learner.
3. **Reinforcement Learning:** Reinforcement learning is different in that it is used in situations that may be modelled as an action-reward pair. The training data in the case of reinforcement learning may not be labelled in themselves, but a feedback mechanism may exist that could indicate to the learner how well it is performing. Reinforcement learning is extensively used in robotics.

FIELDS CLOSELY RELATED TO AI

At the present time, there are many terms in use that are related to AI, are distinct from it, but are sometimes used interchangeably. In order to clarify the definition of these 'buzzwords', here are some of the terms that are floating around the conversations surrounding AI:

1. **Artificial Intelligence:** In today's age, AI seems to reflect more of a general idea, which is manifested by various other fields such as machine learning, and deep learning. With time, new ideas from statistics, information theory, graph theory, etc. have been incorporated into AI-related studies, thus giving rise to various methods which are studied specifically.
2. **Machine Learning and Pattern Recognition (PR):** These two areas include the study of many different techniques,

which include Bayesian learning, tree-based learning, discriminant analysis, support vector machines, etc.[3]. These two fields reflect almost the same set of ideas, with pattern recognition laying a little more emphasis on feature engineering and data acquisition.

3. **Data Mining and Big Data:** Data mining is primarily concerned with the filtering and analysis of data, and the search for relevant patterns. The inference achieved from the data is often used in domain-specific contexts (such as business analytics, automated health diagnostics, etc.). The reason it is treated a little differently from traditional ML and PR is that ML is usually used to denote the field that works on the theory and principles, whereas data mining is often an application of various machine learning techniques to find hidden information from features in data. *Big Data* is more of a buzzword than a technical term, which more or less points to the handling of large volumes of data. Often, the term refers also to the architectural requirements to analyze large volumes of data: usually, frameworks such as Hadoop or Spark are used. *Large volumes of data* refer to many millions to billions to trillions of data samples which need to be analyzed! An example of a large dataset is a dataset of the properties of all the discovered galaxies in our universe (about 2 trillion galaxies), which could possibly be about a few hundred petabytes in size. To gain an appreciation for large volumes of data, if a small cluster of servers (a ballpark figure of 10 servers with 16 cores of 2.2GHz each and almost 50GB of RAM) could perform a certain analysis on this dataset in about 4hrs, a personal computer (dual core, 2.2GHz each core, and 4GB RAM) would probably immediately crash trying to analyze about 1/10000th of the data! Note that this example isn not actual experimental evidence (storage of such large volumes of data on a personal computer would be infeasible, and it's analysis would be nearly impossible), but is representative of the complexity involved in analyzing large volumes of data.
4. **Deep Learning:** This is a new and extremely active area of research within machine learning, drawing inspiration from biological neural systems[5]. The set of techniques encapsulated by deep learning include algorithms which work in multiple layers. The techniques encompassed by deep learning include deep neural networks, recurrent neural networks, etc.
5. **Natural Language Processing and Speech Processing:** These fields have contributed immensely to the automation of mundane tasks like interpretation and answering of questions, generation or summarization of text etc. These enable the machine to read text, hear speech, interpret it based on context, measure sentiment and determine valuable information from it.

INFLUENCE OF POP-CULTURE REPRESENTATIONS AND THE TERM ARTIFICIAL INTELLIGENCE'

Now, on to the issues of the public perception of AI and its potential role in our lives.

Perhaps the first reason for the vilification of AI is its name. Due to the influence of popular culture, many associate the term 'AI' with a humanoid machine: one made of metal, the external anatomy of which is very similar to that of humans, the internal anatomy of which is made up wholly of circuits and chips; which has wide agency in the world and is *taking the role of a human being.* The truth, however, is far from this. Most of AI sits in computers, focused on an extremely narrow task. The large corporations such as IBM, Google, etc. run test AI on huge infrastructures, whereas in an academic lab, the testing might be done on a smaller cluster of server computers. Some notable exceptions are Sofia by Hanson Robotics and Talos by Pal-Robotics, which reside in humanoid bodies. As expected, their similarity to the sci-fi trope has garnered outsized coverage from media sources, which further supports the tropes.

Perhaps the term 'Artificial Intelligence' is beyond salvaging at this point, and a less polarizing term for this technological umbrella (at least for a very wide range of use cases) such as 'Assistive Intelligence' should be adopted. True that AI itself began as a series of experiments trying to emulate human reasoning, but today, despite our advances in infrastructure, it is largely being used to solve very specific tasks, rather than emulate the full functioning of a human being. Additionally, even within this narrow scope, AI routinely runs into difficulties. Some errors could stem from inappropriate usage. Some errors might be a result of inadequate building up of the systems. All the AI-based products that are available today are prone to errors: we all occasionally stumble upon a roundabout route in Google Maps or a misinterpretation by our smartphone's voice assistants (which is indeed one of the dangers of AI as elaborated later). Thus, we can see the current course of AI development is on a very different track than the Hollywood depictions would lead you to believe.

Now that we better understand the current scope and applications of AI, we can turn our attention to considering the ethics of its use. Responsibly applying AI to tasks involves both carefully defining the problems that it addresses, and ensuring that the algorithm is not a black box to the people who are affected by the system. The next section will discuss the importance of these moral considerations, even as the field experiences a period of massive growth.

CASE STUDY: TRANSPARENCY IN HIRING ALGORITHMS

In Cathy O'Neil's alarmingly titled book, *Weapons Of Math Destruction,* she includes the story of Kyle Behm. Kyle was a high achieving student from Vanderbilt University who applied for a job at a department store to earn some extra money in his free time. The notion that he might be rejected for this minimum wage job never crossed his mind, but he never got a call back.

Kyle would later find that he had failed a 'personality test' developed by Kronos Incorporated (a workforce management company) that was part of the recruitment process for the store. Kronos' algorithm had flagged his application based on his answers. Further investigation found that the test could violate the Americans with Disabilities Act of 1990 if courts determine that it qualifies as a 'medical exam'; the case is currently ongoing.

It is pertinent that Kyle Behm only learned why he had been snubbed because:

1. He had a connection within Kroger, the departmental store, who told him he had been flagged for the personality test component of the application.
2. He recognized the test as being similar to a medical test he had taken at an earlier time (the common "five factor model" test) at a hospital, and brought this to the attention of his father.
3. His father was an attorney, and had the resources to investigate the legality of the test and bring a class-action case before the courts.

In this instance, Kyle only lost the chance to make some extra spending money. For thousands of people around the world, not getting a call back from a potential employer with no explanation is a common occurrence, and consequences can be truly disastrous. Livelihoods can be threatened by a specialized algorithm working behind the scenes, and most victims do not have inside connections at the job, or a family with the resources and knowledge to fight the decision. In addition to recruitment tests, algorithms also work to determine insurance rates, no-fly lists, fines in court, and to generate performance reports for teachers at public schools. Admittedly, these are not the edgiest applications of AI and machine learning, but they are the most immediately impactful for an everyday person, and they often fly below the radar due to lack of transparency.

Aside from the problems that arise due to lack of transparency, two further categories of issues raised by AI are elaborated below:

1. **Hard-coding human bias:** For a mathematical model to be consistent is easy - for it to be fair is hard. In many cases, biases such as those related to sex, race, etc.[6] are *learned* by algorithms. Such biases need to be carefully evaluated in order for a system having real-world consequences to be fair and just. Similarly, such biases exist in many domain-specific contexts where the selection of a certain type of decision leads to higher statistical accuracy but does not imply fairness.
2. Potential criminal applications: Like any technology or tool, the power of AI can be harnessed in a variety of destructive ways. In recent news, one specific concern that has arisen is that of 'deepfakes'. Supasorn Suwajanakorn presents cases of *fake videos* in his TED Talk [7] that are almost impossible for humans to detect. Other instances of learning of human tasks can also have dangerous consequences, such as if AI is integrated into weapons technology by rogue actors.

For cases such as these, it is extremely important to have ethical guidelines laid out about the usage and deployment of intelligent systems. The issues are not particularly easy to address, especially due to the easy accessibility of AI. With time, scientists developing such technologies should be fully aware of their implications; usage of AI should become increasingly democratic.

An analogy can be drawn between the above points and scientometrics. Oftentimes, the systems for ranking different authors, journals, papers, etc. suffer from inherent vulnerabilities.[8] These systems can be considered to be a form of AI in themselves. In today's world, a lot depends on ranks and metrics such as H-index and the impact factor of journals in professional circles - employment in universities, research jobs in corporations, and the chances of securing funds for projects. To what extent should the numbers game be trusted in gauging the capabilities of researchers? Understanding the working of such metrics is crucial to fairness in the scientific community. In recent times, it has been shown that a lot of the metrics in use today should be used with thorough consideration instead of being used as a single indicator[8,9].

With time, it is expected that such metrics will become more sophisticated and *smart*.[10] Hence it is important to keep in mind the implications of AI and automation.

CASE STUDY: QUANTIFYING INTERNATIONALITY OF SCIENTIFIC JOURNALS

A subproblem in scientometrics is the quantification of internationality of journals. There are thousands of journals across all areas that are published by various countries. The question regarding the internationality of a journal aims to uncover the global impact that a journal has in its domain.

A journal may claim to be an international journal by a set of rules of thumb. The most basic of these are: having an ISSN number, having editorial members from more than one country, and having paper contributions from more than one country. In a naive way, while these rules of thumb could potentially indicate that a journal is international by the mere fact that people from multiple countries are involved in making sure that the journal is delivering articles, judging *how international* a journal is, is a challenging problem.

At the root of the complexity of solving this problem is the lack of structured data that can be taken as a ground truth or training set. If the above mentioned rules of thumb are to be used as features in an intelligent system (either categorical *yes/no* type features, or as continuous valued features indicating the number of people involved internationally), then these numbers may be easily manipulated by a journal with the sole purpose of getting the system to rate it as highly international.

Provided that reliable scientometric data in this domain can be collected over time, one possible solution to the problem of internationality may be to use unsupervised learning methods to discover clusters of journals that are similar in their international impact. Such an exercise would not only serve as a useful set of methods that could optimally find categories of internationality based on information that is potentially hidden in the data, but would also provide us with useful insights on what causes different categories to arise in the first place. Additionally, a cluster analysis could also provide us with evidence of the distribution of attributes corresponding to each underlying category of internationality.

Over and above, AI is not all about classification or clustering: modelling and optimization are at the heart of machine learning. The point of applying machine learning in scientometrics would be to enable us to solve interesting problems within the domain that have strong inferences towards some aspect of quality of journals, publications or authors, and journal internationality is one example for the same. Regardless of the work that is carried out, it is important to gather reliable data over time so that we may be enabled to identify new and interesting problems and be able to address them appropriately using modeling based on machine learning or otherwise.

THE FUTURE OF AI AND ETHICAL CONSIDERATIONS

As AI moves from niche applications to mainstream areas of science, government, policy making, etc. (a lot of this is already in progress), it is important that we leave space for human interpretation and intervention. In his TED talk on AI, Shyam Sankar talks about the importance of human-computer coordination[12]. In a nutshell, he discusses how AI could be used to facilitate and complement the skills of humans rather than remove humans from the equation entirely. Like any other paradigm-shifting technology, there are significant ethical concerns regarding AI. The strides that AI is making should be accompanied by public awareness, as well as regulation where necessary. These policies should ensure the transparency of the AI systems that will surely be widespread in the coming years.

The reader will note that despite all the advancements and shortcomings we have discussed here, we have steered clear of the issue of the AI singularity - the day that AI surpasses humanity and becomes capable of improving itself, launching itself into a self-improvement cycle that results in an unfathomable level of intelligence. Experts in the field are split on the likelihood of this 'intelligence explosion' theory, with a 2017 survey showing half the responses to be in the 'unlikely' or 'very unlikely' categories, and a considerable proportion responding with 'about even'[13].

As discussions on the singularity are largely uncertain and often steer into quite philosophical territory about the fundamental definition of intelligence, the authors believe that the pressing issues regarding AI are the ones that are closer on the horizon: ensuring that, as AI systems scale, they are transparent and focus on the coordination of man and machine should be the primary goal.

REFERENCES

1. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons. 2019;62(1):15-25. Available from: https://doi.org/10.1016/j. bushor.2018.08.004.
2. Jackson P. Introduction To Expert Systems (3rd Edition). Addison-Wesley; 1998.
3. Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. New York: Wiley; 2001.
4. Boyd S, Vandenberghe L. Convex Optimization. New York, NY, USA: Cambridge University Press; 2004.
5. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. http://www. deeplearningbook.org.
6. Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In: NIPS; 2016.
7. Suwajanakorn S; 2018. Available from: https://www.ted.com/talks/supasorn_suwajanakorn_fake_videos_of_ reaLpeople_and_how_to_spot_them.
8. Chang CL, McAleer M, Oxley L. Coercive journal self citations, impact factor, Journal Influence and Article Influence. Mathematics and Computers in Simulation. 2013 jul;93:190-197. Available from: https://doi.org/10.1016/jj.matcom.2013.04.006.
9. Das AK, Mishra S. Genesis of Altmetrics or Article-level Metrics for Measuring Efficacy of Scholarly Communications: Current Perspectives. 2014;.
10. Ginde G, Saha S, Mathur A, Venkatagiri S, Vadakkepat S, Narasimhamurthy A, *et al.* ScientoBASE: a framework and model for computing scholastic indicators of non- local influence of journals via native data acquisition algorithms. Scientometrics. 2016;108(3):1479-1529.
11. Ginde G, Saha S, Mathur A, Venkatagiri S, Vadakkepat S, Narasimhamurthy A, *et al.* ScientoBASE: a framework and model for computing scholastic indicators of non- local influence of journals via native data acquisition algorithms. Scientometrics. 2016 Jun;108(3):1479-1529. Available from: https://doi.org/10.1007/s11192-016-2006-2.
12. Sankar S; 2012. Available from: https://www.ted.com/talks/shyam_sankar_the_rise_of_ human_computer_cooperation.
13. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O. Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts. Journal of Artificial Intelligence Research. 2018 jul;62:729-754. Available from: https://doi.org/10.1613/jair.1. 11222.