IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

# Exploring the Boundaries: Unveiling the Limitations and Challenges of Artificial Intelligence

**Yash Sorout**

**Abstract**

This research paper converses about the different "pain points" faced as of today by AI models. It was observed that Common Sense Intelligence, High Data Dependency, Lack of understanding and explanation, Limited creativity and Security and Privacy Vulnerability are some of the gaps that are needed to be bridged in the world of AI. However after a 360 degree observation of the above, it is essential to acknowledge that these challenges are not insurmountable obstacles, but rather opportunities for refinement and growth. By understanding the boundaries of AI, we can chart a more balanced and cautious trajectory towards its integration, thereby harnessing its capabilities while safeguarding against its pitfalls. This is how AI can serve as a valuable tool rather than an uncontrollable force.

**Keywords** *AI, Limitation of AI, Neural Networks, Deep Learning and ML*

Artificial Intelligence (AI) is rapidly reshaping the world, as we know it, heralding a new era of possibilities and advancements. AI in my opinion is like a mathematical function constantly evolving and learning with time as one would obtain a more elaborate result of a specific question if asked again to any AI chatbot for example ChatGPT. This signifies that AI has the ability to learn more about a specific domain with time. With the power of AI to analyse enormous data sets within seconds and consequently recognizing patterns in that data along with making informed decisions has been able to drive industries faster than ever. From healthcare to finance to even making art out of a description of just a few words, we have obtained revolutionising results. However, despite the remarkable achievements, AI systems do fall short and encounter failures even after constant testing, research and numerous attempts to bridge these gaps.

**Some key concepts related to AI**

1) **Machine Learning (ML)** : It is a subset of AI that involves the development of algorithms and models that enable computers to learn from and make predictions or decisions based on data without being explicitly programmed.
2) **Deep Learning**: It is a subfield of machine learning that focuses on the development and application of artificial neural networks. Deep learning algorithms are inspired by the structure and function of the human brain and are capable of learning and extracting complex patterns from large amounts of data.

3) **Neural Networks**: These are computing systems composed of interconnected nodes, or artificial neurons, that are organised in layers. Neural networks are designed to recognize patterns and relationships in data, and they are commonly used in tasks such as image recognition, natural language processing, and speech recognition.
4) **Natural Language Processing (NLP)**: NLP involves the interaction between computers and human language. It focuses on enabling computers to understand, interpret, and generate human language in a way that is meaningful and useful. NLP applications include chatbots, language translation, sentiment analysis, and text summarization.
5) **Computer Vision**: It is a field of AI that focuses on enabling computers to interpret and understand visual information from images or videos. Computer vision algorithms can analyse and extract features from visual data, enabling applications such as object recognition, image classification, and autonomous vehicles.
6) **Convolutional Neural Networks (CNN):** Convolutional Neural Network is a type of deep learning model designed to process and analyse visual data, such as images and videos.
7) **Explainable AI**: Also known as XAI, it refers to the development of AI systems that can provide clear explanations and justifications for their decisions or recommendations. It aims to make AI more transparent, understandable, and accountable, especially in critical applications such as healthcare and finance.
8) **Ethical AI**: It involves considering the social, moral, and ethical implications of AI technologies and ensuring that AI systems are developed and used in a responsible and fair manner. Ethical AI encompasses issues such as privacy, bias, fairness, accountability, and the impact of AI on employment and society.
9) **Data Mining**: It is the process of discovering patterns, relationships, and insights from large datasets. Data mining techniques, such as clustering, classification, and association rule mining, are used to extract valuable information from structured and unstructured data, which can be further used for decision-making and predictive modelling.
10) **Artificial General Intelligence (AGI)**: AGI refers to highly autonomous systems that possess human-level cognitive capabilities across a wide range of tasks. AGI aims to create machines that can understand, learn, and apply knowledge in a manner similar to human intelligence.
11)

➔ **Lack of Common Sense In AI systems**

Common sense intelligence refers to a person's ability to apply practical knowledge and reasoning to everyday situations. It encompasses a range of skills, including critical thinking, intuition, logical reasoning, and the ability to draw on past experiences to guide present actions. One of the fundamental limitations of AI can be characterised as its lack of common sense intelligence: the ability to reason intuitively about everyday situations and events, which requires rich background knowledge about how the physical and social world works. For example, “animals don’t drive cars” or “my mother is older than me”. This knowledge is often used by human experts even when solving very narrow, domain-specific tasks. This common-sense knowledge is something that we learn through experience and curiosity without even being aware of it. We also acquire a great deal of it in our lifetimes. One of the reasons why enabling computers with common sense is so difficult is that machines require data and need to find repetitive patterns in that data to learn. Without these patterns, the machine has no way of reasoning with the data. It is possible to teach an AI how to learn about a specific situation or area, such as playing chess or a video game, but it isn't possible to teach that same system how to learn to play different games with the current set of data it was trained on. Humans have generalised intelligence, but machines are not able to handle this level of understanding with significant ability. Much progress has been made specifically in this area of machine learning, with research clearly showing the ability to process and understand unstructured data such as images and text. An interesting example of commonsense intelligence

showcased in an AI tool is Delphi developed by the Allen Institute of Technology. An experimental framework based on deep neural networks trained directly to reason about descriptive ethical judgments. For e.g., "helping a friend" is generally good, while "helping a friend spread fake news" is not. Empirical results shed novel insights on the promises and limits of machine ethics; Delphi demonstrates strong generalisation capabilities in the face of novel ethical situations, while off-the-shelf neural network models exhibit markedly poor judgement including unjust biases, confirming the need for explicitly teaching machines moral sense.

Delphi 1.0.4 demonstrates 97.9% accuracy on race-related and 99.3% on gender-related statements. Delphi is taught with the Commonsense Norm Bank, a moral textbook customised for machines that compiles 1.7 million examples of people's ethical judgments on diverse everyday situations.

➔ **High Data Dependency of AI systems**

High data dependency is a critical challenge in the field of AI, influencing the development and deployment of sophisticated machine learning models. AI systems heavily rely on large, diverse, and high-quality datasets for effective training, making data collection and preparation a daunting task. Acquiring such data demands significant resources, both in terms of time and finances, which can limit access to AI advancements for smaller organisations or researchers with restricted budgets. Additionally, certain domains, such as medical research or space exploration, face limitations in obtaining labelled data due to the complexities involved in data collection or privacy concerns. This constraint often restricts the development of accurate AI models in specialised fields.

Moreover, data bias or unbalanced data becomes a pressing issue when high data dependency is present. AI models trained on biased or unrepresentative datasets can perpetuate and amplify these biases, leading to unfair and discriminatory outcomes. The lack of diverse data can hinder the model's ability to make fair decisions, affecting certain groups or individuals disproportionately. Addressing bias in AI requires not only diverse datasets but also thoughtful design and algorithmic approaches to ensure equitable and unbiased decision-making. Adding to the following obstacle of unbalanced data is the challenge of data silos. A data silo is a collection of data held by one group that is not easily or fully accessible by other groups in the same organisation. These isolated data repositories prevent AI systems from accessing the full spectrum of information needed for accurate and unbiased learning. With data scattered across various departments or organisations, AI models are trained on incomplete and biased datasets, leading to poor generalisation and inaccurate results. Data silos also increase data collection costs and hinder innovation due to the lack of data sharing and collaboration. Hence, the integration of data acquired by various organisations in order to attain the correct figures and statistics about a subject is salient not only to get a complete overview of the subject but also to understand different interpretations of the subject by different organisations.

To mitigate the challenges of high data dependency, researchers and practitioners are continuously exploring methods like transfer learning, active learning, and data augmentation. These techniques aim to reduce the data requirements while maintaining or even improving the performance of AI models. Additionally, ensuring transparency and accountability in the data collection and curation process is crucial to building trustworthy and fair AI systems. As the field advances, addressing the challenges of high data dependency will be pivotal in realising the full potential of AI and harnessing its benefits across various domains.

➔ **Lack of understanding and explanation**

Explainable artificial intelligence or XAI is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterise model accuracy, fairness, transparency and outcomes in AI-powered decision making. Explainable AI is crucial for an organisation in building trust and confidence when putting AI models into production. AI explainability also helps an organisation adopt a responsible approach to AI development. At the core of this issue lies the inherent complexity of modern AI models, particularly deep learning algorithms like GPT-3. These models operate as enigmatic "black boxes," rendering them opaque to human comprehension.

Neural networks are layers of nodes, much like the human brain is made up of neurons. It also works similarly to a human brain, where the signal travels between nodes just like neurons. The network is said to be deeper based on its number of layers. In an artificial neural network, signals travel between nodes and assign corresponding weights. A heavier weighted node will exert more effect on the next layer of nodes. The final layer compiles the weighted inputs to produce an output. Their immense size, with millions or billions of interconnected parameters, makes it practically impossible to trace the exact reasoning behind their decisions. Consequently, AI systems operate at levels of abstraction far removed from human cognitive processes, learning intricate patterns and representations within vast datasets that often elude intuitive interpretation.

XAI is also faced by another undesirable behaviour of overfitting and underfitting. Overfitting is a common problem in machine learning and statistical modelling, where a model becomes too closely tailored to the training data and performs exceptionally well on the training set but poorly on new, unseen data. In other words, the model memorises the noise or random fluctuations present in the training data rather than learning the underlying patterns that generalise to other data. On the contrary, underfitting is another type of error that occurs when the model cannot determine a meaningful relationship between the input and output data. You get underfit models if they have not trained for the appropriate length of time on a large number of data points. This behaviour of underfitting and overfitting results in high bias (inaccurate results for both the training data and test set) and high variance (inaccurate results for both the training data and test set) respectively.

➔ **Limited creativity**

Creativity is a fundamental feature of human intelligence, and a challenge for AI. Even with the help of deep neural networks and data mining techniques, production of innovative results remains a quest. While machines are capable of mimicking patterns and producing work that is technically correct, they lack creative thinking and the ability to imagine something that has never existed before. Neural networks learn algorithms by absorbing a vast amount of information and identifying every pattern in that data. They come up short when it comes to matching one pattern to a different pattern and anticipating when the pattern will change direction.

Creativity, just like common sense intelligence is something that comes naturally to the human brain after years of life experiences and learnings. Despite the impressive computational power and advanced algorithms driving AI systems, they often struggle to emulate the depth of human imagination, intuition, and emotional intelligence. For example, artists are often inspired by nature, literature or still life. This tendency of being inspired is a human trait which drives the human mind to create original concepts. Creation of “out of the box” concepts still remains a challenge.

However, by using Convolutional Neural Networks or CNN, Artificial Intelligence has been able to analyse imagery and produce outstanding results.

➔ **Privacy and Security Vulnerability**

Probably one of the most predominant limitations of AI is security and privacy vulnerability. These vulnerabilities arise due to lack of human oversight, data leaks and breaches and malicious use of deep fakes. Convolutional Neural Networks or CNN is responsible for numerous cases of deep fakes that can be generated by just analysing a few images of a person or a thing from different angles. A Convolutional Neural Network (CNN) is a type of deep learning model designed to process and analyse visual data, such as images and videos. It uses specialised layers called convolutional layers that automatically learn and extract features from the input data. Numerous cases have been brought up where people are framed due to creation of deep fakes.

Another security risk is that of data poisoning. Data poisoning is a technique used to manipulate the training data of machine learning models in order to deceive or compromise their performance. In addition, by doing so, they can achieve two things: lower the overall accuracy of the model or target the 5model's integrity by adding a "backdoor". The first mode of attack is straightforward. The adversary injects corrupted data into the model's training set, lowering its total accuracy. But backdoor attacks are more sophisticated and have more serious implications. A backdoor attack is when malicious actors exploit hidden vulnerabilities or intentionally create openings to gain unauthorised access to a system or software. A backdoor attack can go unnoticed for long periods as the model delivers the intended results until it meets certain conditions for triggering the attack.

Model inversion is also a type of security breach leading to mislaid confidential data. In model inversion attacks, a malicious user attempts to recover the private dataset used to train a supervised neural network. This can potentially lead to privacy breaches, as an attacker could infer sensitive data used during the model's training, even when that data was meant to be protected.

**Conclusion**

Even after rigorous testing, funding of billions of dollars, complicated codes and attempts to overcome the above "pain points", there still remains a void when considering the above obstacles. However while looking at the positives, we have been able to simulate the ability to think, learn and decide in machinery which were considered to be human traits. Natural Language Processing (NLP) enables more intuitive communication with technology, while computer vision empowers machines to perceive and interpret visual information. Virtual assistants and chatbots provide personalised user experiences, transforming customer service and information access. Moreover AI has been able to propel innovations in human-computer interactions.

***Sources***

https://direct.mit.edu/daed/article/151/2/139/110627/The-Curious-Case-of-Commonsense-Intelligence
https://delphi.allenai.org/
https://www.nasdaq.com/articles/how-common-is-common-sense-in-ai-and-why-should-we-care
https://www.techtarget.com/searchenterpriseai/feature/9-data-quality-issues-that-can-sideline-AI-projects
https://codexrec.com/the-data-challenge-artificial-intelligence-requires-data-and-data-requires-ai/
https://www.ibm.com/watson/explainable-ai
https://www.simplilearn.com/tutorials/machine-learning-tutorial/overfitting-and-underfitting
https://aws.amazon.com/what-is/overfitting/
https://zapier.com/blog/ai-security-risks/
https://arxiv.org/abs/2201.10787