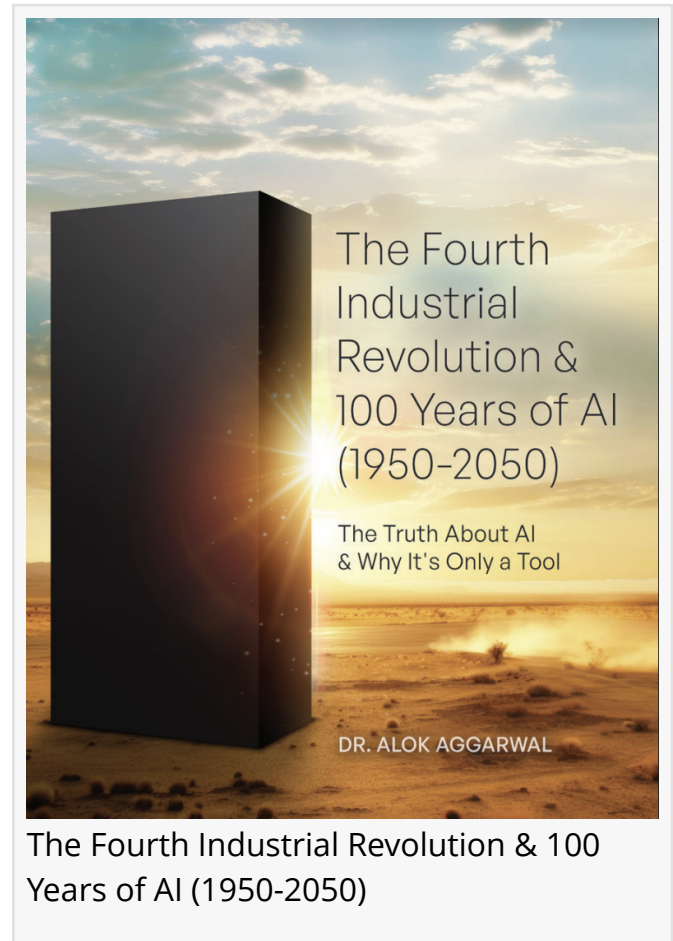# New Book by Alok Aggarwal Underscores the Need for Explainable, Interpretable, and Ethical Artificial Intelligence

The Fourth Industrial Revolution & 100 Years of AI emphasizes the challenges in achieving transparency and predictability

SAN JOSE, CALIFORNIA, USA, March 28, 2024 /EINPresswire.com/ -- A pivotal study conducted in May 2022 delved into the perceptions of 926 randomly chosen patients in the US regarding AI's role in diagnosis and treatment. The findings revealed substantial discomfort among patients when faced with the prospect of receiving a correct diagnosis from a computer program lacking explanatory capabilities. Moreover, the discomfort escalated when AI played a pivotal role in diagnosing cancer, prompting skepticism even among medical professionals. The study sheds light on the existing apprehensions surrounding AI in healthcare, resonating with the reluctance observed among medical doctors who, despite initial promises from IBM Watson, distrusted the system due to a lack of transparency and explainability. Today, we briefly discuss the thirteenth chapter of Alok Aggarwal's new book, "[The Fourth Industrial Revolution & 100 Years of AI (1950-2050)](#)."

The Fourth Industrial Revolution & 100 Years of AI (1950-2050)

The chapter titled, "Explainable, Interpretable, Causal, Fair, and Ethical AI?" underscores the need for Explainable AI (XAI), emphasizing the challenges in achieving transparency and predictability. This chapter provides the following take-aways:

• Explainability: The definition of explainable AI models is somewhat nebulous, but it includes the models being able to provide justification regarding their outputs, provide knowledge especially when they provide counterintuitive results, point out their deficiencies (so that they can be improved), and help in subject matter experts having more control (so that SMEs can

improve these models if required).

• Interpretability: An Interpretable AI model is one that allows quantitative understanding of how one or more features impact the output of the model.

• Contemporary AI systems are neither explainable nor interpretable: Due to the nature of these systems, they rely on non-linear algebra where all inputs get mixed up inextricably, thereby making it extremely hard for researchers to explain the workings of these systems.

• Non-linear algebra underlying AI algorithms makes them unexplainable and non-interpretable: Good interpretation of the working of modern AI systems is also hard because they use non-linear algorithms: Although a few models (e.g., Linear Regression, General Linear Models, Decision Trees) can be interpreted well, they assume that all features are independent of each other, which usually doesn't happen in the real world.

• AI systems are therefore not trustworthy: Since contemporary AI systems are not explainable or interpretable, humans do not trust them. Trust develops over time, but predictability and transparency can help. Predictability requires that these systems are robust with respect to perturbations in data, transparency requires that these AI systems should be explainable.

• Tradeoff often occurs between explainability and accuracy: Another unfortunate aspect is the tradeoff between explainable AI models and their accuracy – in most modern cases, the more explainable models (e.g., linear regression) are less accurate than the less explainable ones (e.g., Deep Learning Networks).

• AI systems do not provide causation: Causal models can provide "cause and effect," i.e., provide the influence by which one or more causes – which could be events, processes, states, or objects – contribute to the effect (i.e., output), and how much each cause contributes to the effect.

• Fair and ethical AI often depends upon the context and human belief systems: Because fairness and ethics are qualitative and are ill-defined in human society, these characteristics will be even harder to define or achieve for AI. Also, just like the regulations regarding data, statutes related to fairness and ethics will vary from one society to another.

• The above-mentioned characteristics are not always required: Even though explainable, interpretable, causal, fair, and ethical AI systems are preferable for many applications and use cases, these characteristics are not always required. In fact, approximately two thirds of the use cases provided in this book are not constrained by such limitations. Hence, progress in developing better AI systems will continue with full vigor even if we are unable to make progress in overcoming these limitations.

Overall, the book, "The Fourth Industrial Revolution & 100 Years of AI (1950-2050)" provides a

concise yet comprehensive exploration of AI, covering its origins, evolutionary trajectory, and its potential ubiquity during the next 27 years. Beginning with an introduction to the fundamental concepts of AI, subsequent chapters delve into its transformative journey with an in-depth analysis of achievements of AI, with a special focus on the potential for job loss and gain. The latter portions of the book examine the limitations of AI, the pivotal role of data in enabling accurate AI systems, and the concept of "good" AI systems. It concludes by contemplating the future of AI, addressing the limitations of classical computing, and exploring alternative technologies (such as Quantum. Photonics, Graphene, and Neuromorphic computing) for ongoing advancements in the field. This book is now available in bookstores and online retailers in Kindle, paperback, and hard cover formats.

About the Author: Dr. Aggarwal is the founder, CEO, and Chief Data Scientist of Scry AI, which provides innovative AI-based products, solutions, and services to enterprises across the globe. Before starting Scry AI, he co-founded Evalueserve (www.evalueserve.com) which provides research and analytics services worldwide. He received his Ph. D. from Johns Hopkins University and worked at IBM's T. J. Watson Research Center during 1984 and 2000. He has written more than 120 research articles and has been granted eight patents. For more information, please visit: www.scryai.com

Alok Aggarwal
Scry AI, Inc.
+1 914-980-4717
email us here
Visit us on social media:
Facebook
Twitter
LinkedIn
Instagram
YouTube

This press release can be viewed online at: https://www.einpresswire.com/article/698594071