# BEYOND THE HYPE: ANDREW RUSSEL ON THE IMPORTANCE OF REPAIR AND CARE

**John Suganya. M PSGCAS**

## INTRODUCTION

Technologies enable the enhancement of human performance in physical, emotional, and cognitive dimensions. The primary advantage to businesses in enhancing humans with technology is to provide a more effective workforce. For instance, imagine if all workers could use wearable technology to answer any product or service question or call up any business data at will. The capacity to enhance productivity, sell more, or serve customers more will enhance drastically. Businesses willing to engage with these technologies should focus on bioacoustics sensing, quantified self, 3D bio printing, brain-computer interface, human augmentation, speech-to-speech translation, wearable user interfaces, augmented reality, and gesture control. The opportunities for machines to displace humans are obvious: hazardous work, simpler but costly-to-do jobs and routine tasks.

The primary advantage is to have machines do the work of humans is enhanced productivity, reduced risk of harm to humans, and occasionally improved quality work or answers. For instance, an extremely competent virtual customer service rep could handle the numerous simple questions from customers and replace most of the customer service reps 'volume' and to work with the latest information. Businesses can seek some of these indicative technologies for innovations on how humans may be replaced by machines: volumetric and holographic displays, self-driving cars, mobile robots, and virtual assistants. . When machines become intelligent and begin to automate more tasks for humans, humans will have to place their trust in the machines and feel secure. The Internet of things technologies will give one greater visibility into machines and the environment they are running in. For instance, IBM's Watson offers 'confidence' ratings on answers it gives to human beings while Baxter depicts a puzzled look on its display when it doesn't understand what to do. MIT is also developing Kismet, which is a robot that perceives social signals from visual and sound sensors, and responds with facial expressions indicating comprehension. Such kinds of technology play a very significant role in enabling humans and machines to collaborate. Internet of Things, machine-to-machine communications services, mesh networks: sensor and activity streams are some of the features of the 2013 Hype Cycle. The boom in big data, analytics, and cognitive computing methodologies will give decision-making support and automation to humans and sense-making and intelligence to machines. All these technologies can be utilized to make humans and things smart. NLQA technology can enhance a virtual customer service rep. NLQA can also be employed by physicians to search enormous quantities of medical journals and clinical trials to assist in diagnosing an illness or select a proper course of treatment. These support technologies are pillars both for machines and humans as we evolve towards a digital destiny and businesses must take into consideration quantum computing, prescriptive analytics, neurobusiness, NLQA, big data, complex event processing, in-memory DBMS, cloud computing, in-memory analytics, and predictive analytics.

AI plays various roles in crime reduction, AI in criminology performs more implications by defining advancements and limitations. LSTM networks predict crime occurrences within 75 to 90% proves AI’s ability to improve public security in urban settings such as a major city by leveraging spatial and

temporal data analysis. This artificial intelligence is used in justice making to predict, detect and also to prevent criminal activities. There are numerous issues in crime preventions such as new cybercrime strategies, resource limitations, cross-border offenses and also prejudice and privacy issues but AI can identify possible crime hotspots, offers real-time alerts to enable officers to intervene ahead of where crime takes place.

In digital behaviour analysis AI-based OSINT solutions may analyse social media information to identify suspicious signs and also links. AI has highly improved cybercrime preventions through analysing network traffic and identifying malicious activities.

Machine learning model have significantly improved fraud detections mostly used in financial institutions reporting accuracy rates of 95%, it also enhances investigate efficiency but all this advancement comes with a significant risk. Policymakers and law enforcement officials are exposed to between capitalizing on innovations and also responding to ethical concerns like bias, privacy and also in terms of resources.

## THE FUTURE PROGRESS OF AI

Artificial Intelligence is a wonderful integration of science and engineering that makes machines clever and human-like thinkers. Reading about the recent experiments, it is not overstatement to say that the target of scientists is to go beyond human-level intelligence as machines are able to think and make decisions by themselves and can work endlessly to accomplish their duties. There are numerous best artificial intelligence businesses in continuous efforts to apply AI-powered mobile applications and software to develop personalized solutions for any vertical industry under the sun. It is no hidden fact that AI pervades everywhere because it has not stopped at computer programming and science but has advanced into other domains like health, music, art, business, security, and even in the education sector. Most of the mobile apps empowered by the AI are deeply entrenched within all business processes of all industry verticals. It is reported that within the next few years, AI will be contributing to nearly every industry. The future of Artificial Intelligence is with AI researchers and they are going all out with utilizing the technology to develop landmarks in various industries and manufacturing sectors. Developments in artificial intelligence are seen at a never-before pace and IT outsourcing firms are quick to follow the latest trend in order to add it to their practices. However, leading developers would utilize the advantages of new AI tools to develop front and backend in order to customize a product that not only drives sales but also offer an unprecedented user experience ensuring the super success of the product. On the positive side, Russell performs very well, in the initial chapters, of explaining the power and limitations of modern AI systems. Something he mentions with respect to the systems that have defeated world's top chess, Go, and other game players, is that the systems are merely playing 2 player games. Much of that can be solved by brute force look ahead. More sophisticated AI will be required to decide interplay among larger numbers of players. The standout chapter for me is chapter four. He examines some of the early misuses of AI, such as government monitoring and control, robotized weapons, and job loss. The final is where I believe the greatest threat of societal disruption lies, and one that the book doesn't devote enough space to exploring. That trip is another great portion of the book. The author alludes to several individuals who proclaim a super-intelligent AI is still far down the line. He does address that by referring to two items. First, no one has any idea how close or how far it actually is. Second, significant changes need to be contemplated ahead of time. That is simple, but so many individuals overlook challenges until there is no choice but to deal with them. A good deal of the latter half of the book is consumed by what was said in the last paragraph: the super-smart AI. I cannot pick at what that entails and how we can be sure but the question to ask is how do humans control those machines or at least keep from being

controlled by them. That is where all the interesting questions get asked, and nice to read them; but also where the answers aren't really discussed. Much of the second half of the book is taken up with what was mentioned in the previous paragraph: the super-intelligent AI. I can quibble with what that means and how we'd know but an important question to ask is how humans can control those machines or at least prevent them from controlling us. This is where lots of good question are raised, and it's nice to see them; but it's also where the answers aren't actually discussed. A great deal of the second half of the book is spent on what was described in the last paragraph: the super-smart AI. I can quibble with what that means and how we'd know but an important question to ask is how humans can control those machines or at least prevent them from controlling us. This is where a lot of interesting question are asked, and it's nice to see them; but it's also where the answers aren't really addressed. The late Melvin Kranzberg, a professor of the history of technology at Georgia Institute of Technology, formulated his "laws of innovation" over thirty years ago. These laws apparently still apply to a small number of scientists and engineers who have been deeply affected by them. Two of the six "laws" are of special importance to the issue of controlling super intelligent AI. Russell, co-author of one of the most popular AI textbooks Artificial Intelligence: A Modern Approach (now in its fourth edition) has walked a tightrope between the pro-AI tribe and the anti-AI tribe carefully. He acknowledges the potential beneficial applications for AI in society (e.g., advances in scientific study) as well as potential misuse of AI (e.g., automated extortion). Overlooking the possibility for super intelligent AI technology to be a disaster for humanity would be extremely dangerous. In fact, Russell warns that silence on speaking out against possible disastrous consequences will only guarantee more chance of this endgame. Russell is following the precautionary principle of innovation. That principle has actually been utilized politically to stifle progress on technological innovation that would be net-positive for society. Rather, I would describe his style as one of "responsible innovation" that charges all the stakeholders within the system of innovation with a spirit of individual and shared responsibility. Although I will pick minor nits on some of his arguments, I commend his three principles for good AI. We must learn to control our AI technology before it learns to control us. For those who wish to be a part of this new, critical debate on AI, I would strongly suggest reading and considering this highly informative, very important book. Russell takes the precautionary principle of innovation. That principle has been employed for political purposes to suppress progress on technological innovation that would benefit society in the net. Rather, I would describe his methodology as one of "responsible innovation," in which all of the stakeholders in the system of innovation need to own a sense of personal and shared responsibility.

While I might nitpick over some of his lesser points, I applaud his three principles on good AI. We must learn to control our AI technology before it learns to control us. For those who wish to engage in this new, critical discourse surrounding AI, I strongly suggest reading and considering this highly informative and valuable book. Russell follows the principle of caution regarding innovation. That principle has been applied politically to prevent progress towards technological innovation that would benefit society in the long run. While I might nitpick at some of his lesser observations, I welcome his three principles for good AI. We must learn to control our AI technology before it learns to control us. For all those interested in joining this new, critical debate on AI, I would strongly suggest reading and considering this extremely useful and significant book. AI in medicine assists with anything from diagnosing conditions to the creation of new drugs. It scans vast amounts of information to identify patterns so that doctors can diagnose disease earlier and tailor treatment to individuals. While AI enhances efficiency and patient care, we need to address data privacy and bias concerns so that it's applied fairly and ethically. At the end of the day, AI assists medical professionals, complementing what they can do and making healthcare more efficient and accessible. It is not my job to decide if Russell has cracked the problem of human compatibility or not, but it is significant that his suggestion

comes out of the very requirements of AI research itself. Additionally, it's not just a catchphrase: Russell's answer is a mathematics project formulated as a game. To be able to describe how it can work, Russell must bring in some of the mathematical jargon, with special focus on Turing machines, game theory, and computational complexity. By the end of the book even the mathematically uninformed reader will have encountered intelligible introductions to the leading contemporary methods of machine learning. This game theory is also applied to study how prisoner's dilemma can be applied in actual life scenarios to solve crimes. The game theory, particularly the concept of Nash equilibrium, lies at the centre of Russell's solution scheme, exactly because "human compatibility" proves on scrutiny to be a very much flawed concept. In addition, it's not just a catchphrase: Russell's solution is a mathematical project modelled as a game. In order to explain how it does, Russell must bring in some of the maths talk, with special care to Turing machines, game theory, and computational complexity. Russell presents the necessity of creating mathematical models to have AI systems follow a safe path and how this can be used in crime prevention. The models would incorporate a new approach, partly based on three design rules outlined in the book. Upon reading the material, I found myself with an understanding and appreciation for the imperative need for Explainable AI, well beyond ethics and justice. He quotes a figure that there are currently over 300 worldwide AI Ethics groups. This is sad because we likely only need one good worldwide group examining AI Ethics. Without Explainable AI, we can't generate good data's across all domains. The book concerns issues of final control and challenges as AI penetrates deeper and deeper into society. It is a well-written alert to all of us that we must alter course. Russell quotes super references and has plenty of good examples supporting his reasoning.

## CONCLUSION

Since we are at the intersection of current and future Russel has led us in the right direction by persuading us to select the right path judiciously.

It talks about the significance of AI to the humans and framing the discussion around AI. This book evidently describes the effects of AI on the society and also this book becomes a blueprint for all the advancements of AI and also a guideline for humans regarding where we should go. Based on the points of Andrew L. Russell and Stuart Russell, a concise conclusion points out a double challenge for AI in healthcare.

Russell's emphasis on the "ethos of care" implies that although AI can streamline and mechanize, the human factor of empathy in healthcare cannot be emulated. This reiterates AI's function as a tool for assisting, not substituting, human caregivers.

Stuart Russell's writing, in the meantime, cautions against the necessity of "human-compatible" AI that is aligned with our values, emphasizing the essential importance of developing AI systems that are not only intelligent but also provably safe and beneficial. Finally, the effective integration of AI in healthcare and also in the aspect of crime reduction and defence relies on an accurate balance between technological progress and the maintenance of human - oriented values and ethics and thus continues to benefit the society in various ways.

REFERENCES:

1. Stuart Human Compatible - Artificial Intelligence and the problem of control, October 2019 : https://www.amazon.com/Human-Compatible-Artificial-Intelligence-Problem-ebook/dp/B07N5J5FTS

2. Stuart Russell, Peter Norvig, Artificial Intelligence: A Modern Approach, April 2020 : https://www.amazon.com/Artificial-Intelligence-Modern-Approach-4th/dp/0134610997/ref=dp_ob_title_bk

3. Yuval Noah Harari, Homo Deus: A Brief History of Tomorrow, February 2017 : https://www.amazon.com/Homo-Deus-Brief-History-Tomorrow/dp/0062464310

Copyright & License:

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.