How to address shortcomings in AI to build trust

wipro

Amazon successfully used automation and artificial intelligence for order recommendations and fulfillment in their ecommerce platform and warehouses, but when it came to evaluating job candidates, the scoring engine failed[1]. Built-in biases in the real world were amplified in the machine world with recruiters realizing that scoring results for technical positions like software development were not gender-neutral and qualified women candidates were not being recommended. Amazon scrapped the project.

Facebook shut down its AI system meant to learn negotiation skills, after bots created their own language that was nonsensical and difficult to debug[2]. Microsoft had false starts with a chatbot that learned racist and profane terms in interactions with consumers[3].

Artificial Intelligence holds great promise, but the challenges in implementing this technology has far-reaching implications. It has blind spots that are obvious when bot behavior is part of the solution (profanity) but more difficult to address when processes and cultural behavior influence the technology (racial, gender, socio-economic variables).

Why the lack of trust in AI?

Privacy threat
Fear of consumer data privacy breach

Lack of transparency
AI is a black box with incomprehensible complexities

Bias in decisions
Impact of race, gender and socio-economic factors built into data

Intuitive decisions
Based on multiple algorithms and calculations

Lack of trust in AI

In spite of all the technological advances, people still lack complete confidence in automation and predictive insights, and the underlying relationships organizations have with consumers.

A large part of this relationship is built on the data organizations collect from consumers.

Is the data accurate? Is the data private? Is there recourse to correct data?

Google was in the headlines when it shutdown Google+ after a software glitch in the social site holding personal consumer data was found to be susceptible to hacking[4]. There is no evidence that this data was improperly accessed but it

It is clear that organizations embarking on a data-driven journey must understand the basis of trust in systems and data that consumers expect.

caused some PR issues. Bias is a societal issue and because humans program systems, those systems are susceptible to existing bias. And since the data to teach and test systems comes from human action, the bias can be built into the data.The decisions that AI makes is sometimes highly intuitive. Since it undergoes a host of calculations with the data along with variables, it gets difficult for a human to understand the basis of the decision. This is the 'black box' problem with AI. With the lack of transparency, humans are not able to understand what goes on inside AI, which causes anxiety and ultimately distrust.

It is clear that organizations embarking on a data-driven journey must understand the basis of trust in systems and data that consumers expect.

Creating transparency

Research suggests that using AI to augment human decision-making will allow the AI system to learn from human experience and improve trust. Assembling stakeholders from diverse backgrounds can address the challenge of human bias through governance and other research to understand how algorithms make choices.

An effective data governance strategy is essential in order to address error and bias in a data model. This is a key component of an AI system framework since it not only offers a simple way to use the right data but also flags the errors in the data.

A framework for evaluating data fit for AI:

- Involve the data governance team in educating the AI teams on regulatory, data privacy and governance issues.

- Develop a minimum data quality check list for AI that sets the baseline for all AI experiments. Start with regulatory, privacy and governance rules based on an organization's digital ethics.

- Base a data management and governance practice on trust, transparency and diversity of inputs and data sources.

- Build a practice involving data engineers and data scientists in devising how to assess the performance of the models and algorithms, and iterating as needed to refine the inputs to improve performance for systems automation and augmentation in decision making.

The core of data governance is not only about high quality data but also about the legal requirements of handling consumer and personal data combined with an organization's digital ethics. Regulations like GDPR allow consumers to keep a tab on their personal identifiable information (PII), rather than it being floated around without their consent. Digital ethics is the foundation for how algorithms are created, tested, and deployed. Furthermore, well-organized data will increase consumer confidence in a company's brand when consumers understand who owns their data and why.

Making AI more insightful

AI is the 'black box' that crunches a vast amount of data to identify hidden patterns and underlying signals through its complex system, whereas human trust is often based on how people think. It is pertinent to open the black box of machine learning and be translucent if not completely transparent to the user on its working. Some of the analysis performed by AI will inevitably be probabilistic based on incomplete information. It is therefore important that an organization recognize the limitations and explain this to its customers through a compelling user experience.
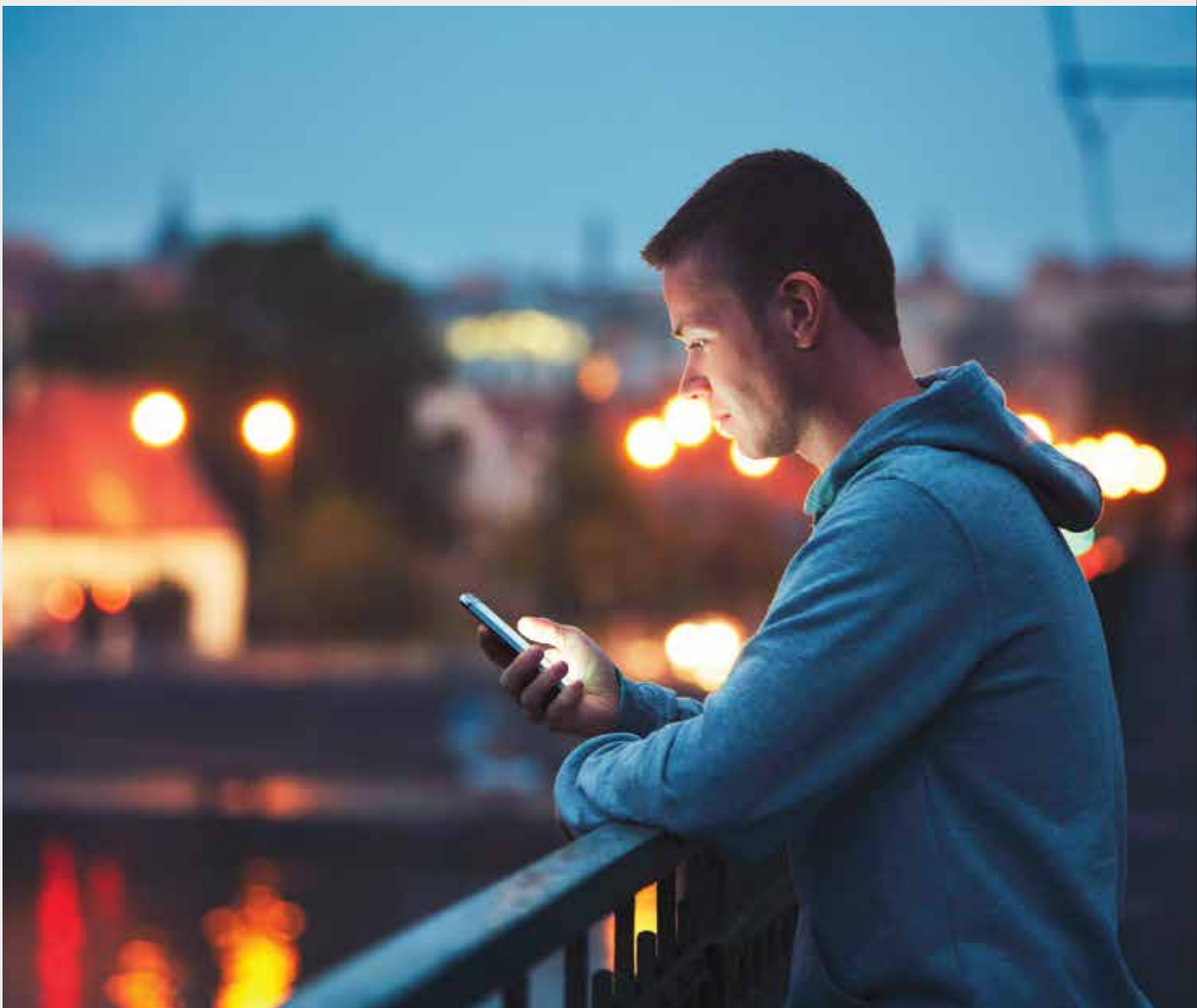
References:

[1] https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

[2] https://techcrunch.com/2017/09/06/the-secret-language-of-chatbots/

[3] https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/

[4] https://www.wsj.com/articles/google-exposed-user-data-feared-repercussions-of-disclosing-to-public-1539017194

About the authors

Alex Soejarto
Head of Strategy – Data, Analytics & AI, Wipro Limited.

Alex is a thought leader in IT services disruption and innovation. He has worked with solution providers in their investment strategies and competitive positioning. He leads the strategy, planning, and marketing teams of the DAAI service line at Wipro.

Roshan Wilson
Consultant, Strategy & Planning – Data, Analytics & AI, Wipro Limited.

Roshan is currently responsible for building value propositions and driving cohesive strategies for Wipro's Data, Analytics & Artificial Intelligence business. He has significant experience in strategy, thought leadership and marketing. Roshan is a Computer Science Engineer and holds an MBA degree from FORE School of Management, New Delhi.

Wipro Limited
Doddakannelli, Sarjapur Road,
Bangalore-560 035,
India

Tel: +91 (80) 2844 0011
Fax: +91 (80) 2844 0256
wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information, please write to us at **info@wipro.com**

Wipro Limited
Doddakannelli, Sarjapur Road,
Bangalore-560 035,
India