

Marketplace Update

March 2019

- **IBM's "Explainable" AI – Building Transparency and Trust into Artificial Intelligence**

By Charles King, Pund-IT, Inc.

Pund-IT, Inc.
Hayward, CA
U.S.A. 94541

Contact:
charles@pund-it.com
Ph: 510-909-0750
www.pund-it.com

IBM's "Explainable" AI – Building Transparency and Trust into Artificial Intelligence

By Charles King, Pund-IT, Inc.

Introduction

Issues of trust seldom arise in discussions about modern computing systems. It's not that hardware and software are perfect. In fact, publications and online forums contain tens of thousands of posts hashing out the relative merits of various PCs, workstations and servers. But those products have been so commonplace for so long that their essential "rightness" as well as the results they provide are hardly ever questioned.

However, that wasn't always the case, and a similar dynamic applies to most all emerging technical and scientific breakthroughs, including commercial artificial intelligence (AI) solutions designed for businesses and other organizations. Considering the inherent complexity of machine learning, neural networks and other AI-related processes, customers' confusion about AI isn't all that surprising. But what can be done to assuage their misgivings and bring AI into the mainstream?

Vendors, including IBM are tackling the problem with solutions designed to make AI processes and results more explainable, understandable and trustworthy. That should satisfy clients' doubts and accelerate the adoption of commercial offerings, but explainable AI also yields other significant benefits. Let's consider why explainable AI is so important and how IBM's innovations are impacting its customers.

The problem of "black box" AI

A lack of clarity or understanding is usually problematic. When it comes to inexplicable artificial intelligence, three potential issues can arise:

1. Most importantly, a lack of transparency leaves users uncertain about the validity and accuracy of results. That is, the essential value of AI projects and processes is undermined.
2. In addition, if AI projects are inexplicable, it's possible that their results might be contaminated by bias or inaccuracies. Call this a problem that you can't really be sure you have.
3. Finally, when AI processes are not explainable, troubleshooting anomalous results is difficult or even impossible. That is, a lack of transparency leaves organizations unable to fix what's broken.

How are AI-focused vendors addressing these issues? Unfortunately, often with fixes that worsen the situation, including "black box" solutions. These purport to deliver all the benefits of AI but fail to provide adequate transparency into how they work, how customers can determine the accuracy of results or how problematic issues can be addressed.

These solutions also encourage perceptions of AI as a mystery whose capabilities can't be understood by mere mortals. In other words, whatever modest benefits "black box" AI may offer, leaving customers in the dark is detrimental to their work and goals.

The benefits of explainable AI

Is there a better way to proceed? Absolutely. How can organizations explain AI successfully? With a holistic approach that addresses several stages of the AI lifecycle:

- Central to making AI projects more explainable is making AI models explainable instead of the black boxes many currently are.
- Organizations must clearly and transparently articulate where and how AI is being used in business processes and for what ends.
- They must also allow for analysis of AI outcomes and provide hooks to alter and override those outcomes as necessary.

To these ends, expert users and managers should employ technologies and solutions that vendors have designed to enhance AI transparency. These methodologies can speed the understanding of AI, which is great. It is also a critical issue for IT, marketing, sales and customer care organizations, especially those in highly regulated industries, such as banking and insurance.

This process can occur organically, as people experience AI and come to understand how it affects the business and them personally, thus impacting the very culture of an organization. Or it can be pursued proactively with the best tools and solutions currently available. Whichever way a company proceeds, people need to keep in mind the vast potential of AI. Why so? Because a time will come when AI is as essential to an organization's success as the business technologies that are commonplace today.

The business benefits of explainable AI

Why is explainable artificial intelligence such an important issue and undertaking? It goes to the practical roots of how organizations do business. If they are to adopt and adapt to AI processes, they need to know that results are accurate. Otherwise, how can they assure customers and partners that AI-impacted decisions are valid and dependable? Consider two examples:

1. In financial services, accurate results are obviously critical for maximizing business outcomes and customer interactions. However, like other businesses in highly regulated industries, banks and other financial organizations must be able to prove that AI-impacted processes comply with government and industry rules or risk significant sanctions and penalties. That's bad enough, but inexplicable AI might also damage client relationships. If customers seeking loans, credit cards or other services are denied by an AI-related system, company officials must be able to explain why that determination was made and how the client might address or correct problematic issues.
2. Global supply chain management is another promising area for AI because the complexity, volume and diversity of supply chain data make it extremely difficult for people to effectively track and adjust for real-time changes in demand. AI can enhance Forecast Value Added (FVA) metrics—learning from past successful and unsuccessful forecasts to help planners make better adjustments. But unless supply chain teams can easily monitor the accuracy of AI models, they can't be certain that systems are really delivering the benefits they promise.

In light of these and other points, it's difficult to see why vendors would develop or customers would consider inexplicable AI solutions.

What IBM is doing to open “black box” AI

IBM is working in numerous areas to develop and deliver explainable AI and advanced analytics solutions. The impetus for the company's efforts was underscored in a recent blog by Ritika Gunnar, VP of IBM's Watson Data and AI organization. “As humans, we're used to the idea that decisions are based on a chain of evidence and logical reasoning anyone can follow. But if an AI system makes recommendations based on different or unknown criteria, it's much more difficult to trust and explain the outcomes.”

Central to the company's efforts is the Watson OpenScale platform that IBM launched in 2018. Designed to “break open the ‘black box’ at the heart” of AI models, Watson OpenScale simplifies AI processes, including detailing how recommendations are being made and automatically detecting and mitigating bias to ensure that fair, trusted outcomes are produced.

IBM is leveraging both existing open source technologies and proprietary algorithms developed at IBM Research to explore, enhance and explicate AI decision-making.

- LIME (Locally Interpretable Model-Agnostic Explanations) is a widely used open source algorithm designed to explain predictions made by AI systems by comparing an explanation to an easily interpretable model.
- Developed by IBM Research, MACEM (Model Agnostic Contrastive Explanations Method) goes well-beyond the capabilities of LIME by identifying both pertinent features that are present in a piece of data and those that are absent, enabling the construction of “contrastive explanations”.

One scenario for contrastive explanations is in banking, where it could be used to analyze loan application data. The system would alert the bank to issues, including poor credit ratings, but it could also spot and highlight missing documents, like an incomplete credit report. The bank could then notify the customer about the reasons for its decision and provide constructive advice.

In essence, solutions that deliver more accurate, transparent and trustworthy AI results, such as IBM Watson OpenScale, can help businesses make better decisions and enhance their services for and relationships with customers.

Final analysis

People are often concerned about new technologies, especially those that are highly complex or difficult to understand. Overcoming those doubts is central to technologies becoming widely trusted and commercially successful. In fact, without fostering understanding of and insights into emerging technologies, it's unlikely that new solutions will find a place among the people and organizations they might otherwise benefit.

By making technologies like artificial intelligence and AI-based solutions and services clearly explainable, vendors can reduce the time required for new offerings to enter the mainstream. That's why explainable AI offerings, like IBM Watson OpenScale are so important.

By breaking open the “black box at the heart of AI” to make processes and results fully explainable, IBM is aiding its customers and partners and furthering its own market strategies. More importantly, IBM’s explainable AI efforts should help establish the essential “rightness” of these solutions as entirely valid and wholly valuable business technologies.

Overall, IBM’s work in explainable AI should improve the mainstream understanding, acceptance and adoption of artificial intelligence among individuals and organizations worldwide.

© 2019 Pund-IT, Inc. All rights reserved.

About Pund-IT, Inc.

Pund-IT™ (www.pund-it.com) emphasizes understanding technology and product evolution and interpreting the effects these changes will have on business customers and the greater IT marketplace. Though Pund-IT provides consulting and other services to technology vendors, the opinions expressed in this commentary are those of the author alone.