IBM UK Newsroom

IBM Takes Major Step in Breaking Open the Black Box of AI

New cloud-based, bias-detection and mitigation controls address need for more transparency in AI decision making

London, UK - 19 Sep 2018: The software service, which automatically detects bias and explains how AI makes decisions – as the decisions are being made – runs on the IBM Cloud, and helps organizations manage AI systems from a wide variety of industry players. IBM Services will also work with businesses to help them harness the new software service. In addition, IBM Research will release into the open source community an AI bias detection and mitigation toolkit, bringing forward tools and education to encourage global collaboration around addressing bias in AI.

"IBM led the industry in establishing Trust and Transparency principles for the development of new AI technologies," said Beth Smith, General Manager of Watson AI at IBM. "It's time to translate principles into practice. We are giving new transparency and control to the businesses who use AI and face the most potential risk from any flawed decision making."

These developments come on the back of new research by IBM's Institute for Business Value, which reveals that while 82 percent of enterprises are considering AI deployments, 60 percent fear liability issues and 63 percent lack the in-house talent to confidently manage the technology.

Visibility into AI decisions

IBM's new Trust and Transparency capabilities on the IBM Cloud work with models built from a wide variety of machine learning frameworks and AI-build environments such as Watson, Tensorflow, SparkML, AWS SageMaker, and AzureML. This means organizations can take advantage of these new controls for most of the popular AI frameworks used by enterprises.

The software service can also be programmed to monitor the unique decision factors of any business workflow, enabling it to be customized to the specific organizational use.

The fully automated software service explains decision-making and detects bias in AI models at runtime – as decisions are being made – capturing potentially unfair outcomes as they occur. Importantly, it also automatically recommends data to add to the model to help mitigate any bias it has detected.

Explanations are provided in easy to understand terms, showing which factors weighted the decision in one direction vs. another, the confidence in the recommendation, and the factors behind that confidence. Also, the records of the model's accuracy, performance and fairness, and the lineage of the AI systems, are easily traced and recalled for customer service, regulatory or compliance reasons – such as GDPR compliance.

All of these capabilities are accessed through visual dashboards, giving business users an unparalleled ability to understand, explain and manage AI-led decisions, and reducing dependency on specialized AI skills.

IBM is also making available new consulting services to help companies design business processes and human-AI interfaces to further minimize the impact of bias in decision making.

Empowering the open source community to build fairer AI

In addition, IBM Research is making available to the open source community the AI Fairness 360 toolkit – a library of novel algorithms, code, and tutorials that will give academics, researchers, and data scientists tools and knowledge to integrate bias detection as they build and deploy machine learning models. While other open-source resources have focused solely on checking for bias in training data, the IBM AI Fairness 360 toolkit created by IBM Research will help check for and mitigate bias in AI models. It invites the global open source community to work together to advance the science and make it easier to address bias in AI. You can read more in a blog here.

Study reveals priorities and hurdles for mainstream AI deployment

According to IBM's just-released study of 5,000 C-Suite executives, the IBM Institute for Business Value AI 2018 Report, there is a significant shift underway in how business leaders look at AI's potential to drive business value and revenue growth.

Among the key findings:

- 82% of enterprises, and 93% of high-performing enterprises, are now considering or moving ahead with AI adoption with a focus on revenue generation.
- 60% fear liability issues and 63% lack the skills to harness AI's potential.
- CEO's perceive the greatest value in
- AI adoption to be in IT, information security, innovation, customer service, and risk management. AI adoption is higher and likely to accelerate faster in more digitized industries like financial services.

About IBM & Artificial Intelligence

A world leader in AI software, services, and technology for business, IBM has deployed Watson AI solutions in thousands of engagements with clients across 20 industries and 80 countries. IBM's Watson AI solutions are widely used in industries, including by seven of the 10 largest automotive companies and 8 of the 10 largest oil and gas companies.

Contact(s) information

Lucy Linthwaite

Mrs +44 020 7021 8911lucy.linthwaite@uk.ibm.com