Driving Value with Explainable AI

The 451 Take

AI technology is becoming increasingly pervasive across all industries; already, more than half of enterprises have adopted AI, and this number continues to rise each year. According to 451 Research's Voice of the Enterprise: AI and Machine Learning Use Cases survey, nearly 29% of enterprises currently have AI in production, and an additional 28% are in the proof-of-concept stage. Estimates suggest that by 2022, over three-quarters of enterprises will have adopted AI.

These numbers come as no surprise, since AI technology can bring many benefits to enterprises, from increasing efficiency in decision-making to modernizing applications and improving business agility. Despite these benefits, the notorious 'black box' nature of AI technology brings inherent risk – risk that grows exponentially as AI scales to production environments. It is imperative that enterprises have visibility into the decisioning process in order to prevent bias, maintain regulatory compliance and improve their AI models.

Importance of Explainability

Source: 451 Research's Voice of the Enterprise: AI and Machine Learning Use Cases 2020

Q. Of the machine learning applications in production, how important is it for their predictions to be explainable to nontechnical users?

Base: Adopted ML in production or proof of concept (n=796)

- Very important: 48%
- Somewhat important: 44%
- Not very important: 7%
- Not at all important: 1%
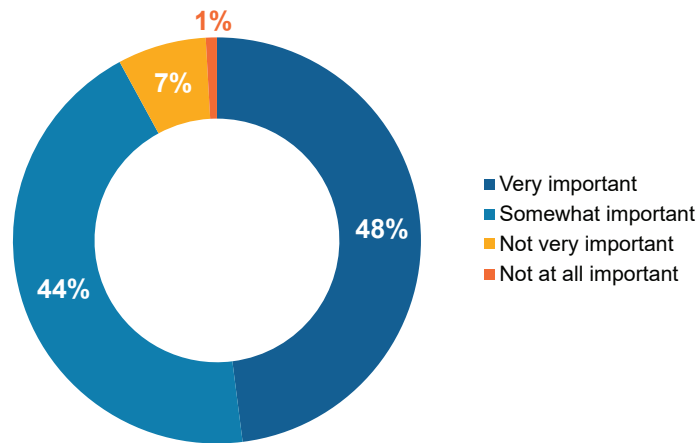
Among businesses surveyed, 92% of AI adopters say explainability is important, 48% of which consider it very important. Despite this, less than half of enterprises have built or purchased tools for explainability. One possible reason for this disparity, beyond the relative nascency of the market, may be the slow pace at which the industry has built stand-alone products or developed integrated features to address this critical problem. Data scientists have long focused on the *what* vs. the *why* in AI models, under the misconception that explainability must be sacrificed for greater accuracy and model performance. But in fact the reverse is true: greater explainability can help eliminate bias and increase depth of analysis, while making models more dynamic and flexible under unpredictable circumstances. Additionally, it frees up time for data scientists to focus on model development.

Addressing transparency and explainability in AI systems is a crucial step for enterprises in fully realizing their AI initiatives. Breaking down the AI 'black box' can lead to significant business value. By understanding the *why* behind AI decisions, enterprises can make improvements in the future that boost their commercial value and build trust among their customers.

Business Impact

VISIBILITY IS ESSENTIAL FOR ENTERPRISES TO EFFICIENTLY UNDERSTAND AND IMPROVE THEIR AI MODELS WHILE PROMOTING BEST BUSINESS PRACTICES. Understanding what goes on 'behind the curtain' not only helps enterprises mitigate risk through bias testing and debug processes, it also promotes deeper and potentially novel insights, which can then be used to further improve model accuracy and business operations.

ENTERPRISES NEED TO ESTABLISH TRANSPARENCY AND CUSTOMER TRUST IN THEIR AI SYSTEMS. AI is employed in use cases ranging from medical diagnosis to financial investing – areas where a small error could lead to severe consequences. As such it is paramount that enterprises are able to ensure the accuracy and security of their AI systems and establish trust with their customers. Understanding model algorithms and decisioning processes is key to ensuring an uncompromised system and maintaining transparency. Transparency is also crucial for ethics considerations and maintaining regulatory compliance. Without visibility, enterprises are unable to test for error or bias, and may face legal or compliance issues – this is particularly crucial in the healthcare and financial sectors.

ENTERPRISES SHOULD PRIORITIZE HUMAN-CENTRIC DEVELOPMENT ENVIRONMENTS THAT INTEGRATE EXPLAINABILITY TOOLING. Enterprises desire efficient development environments. A key element to this is maintaining human-centricity at every stage of the development cycle. When applied retroactively, explainability tools don't offer the same accuracy, visibility or depth of analysis into the AI 'black box' that they would if integrated throughout the development process. AI needs human input to properly interpret the algorithms it's being trained with, and accordingly humans need visibility into their AI to build and adjust these algorithms along the way.

Looking Ahead

AI explainability is already important to the majority of enterprises, and its importance will only grow as use cases become more complex and enterprises scale their AI initiatives. Further adoption of deep learning and the increasing size of ML models means explainability becomes harder but also more essential. Less than half of enterprises currently utilize explainability tools, exposing them to a considerable amount of risk. There is likely to be more regulation of AI technology in general, especially in insurance, and enterprises must be able to explain their models to regulators and nontechnical audiences.

The high demand for explainability tools will undoubtedly lead to more vendors offering new techniques and best business practices for employing explainable AI. The greatest shift likely to be seen going forward is the democratization of AI. Enterprises are made up of many nontechnical individuals and teams that can benefit from AI technology. Explainability tools give enterprises more flexibility to deploy models without dependency on data scientists. Explainability is an essential part of AI strategy; enterprises pursuing AI initiatives only stand to benefit from incorporating explainability tools into their development environments.

H₂O.ai

When laying out all the upsides and mitigated downsides by deploying AI responsibly, organizations may ask "Where do I begin," and h2o.ai/responsible-ai is a good place to start. H2O has the AI expertise and technology to help you innovate responsibly and make getting real results out of your AI that can be explained, understood, and trusted a real possibility.