H2O.ai

Explaining Explainable AI

How to elicit trust and understanding from AI

Can you trust and understand AI?

When considering AI, you may have potential questions or simply want to understand AI better. Trusting and understanding AI is key to a successful AI transformation.

This is especially pertinent in industries with stringent regulations, such as financial services, banking, and healthcare. After all, if the AI gets it wrong, people may be adversely affected.

In this eBook we'll define and explore "Explainable AI"—the concept and the platform, that you can use for more effective, and trustworthy, decision-making in your business.

Regulatory compliance

Regulated industries such as financial services, banking, and healthcare must adhere to stringent requirements. Regulations often require that you can both describe and document your processes in the case of an audit or an inquiry. An organization may need to explain how a customer's credit score was determined, why they might have been denied a loan, or prove why a specific healthcare treatment versus an alternative was decided upon.

What is Explainable AI?

Explainable AI is a relatively new field of AI that attempts to provide explanations and definitions for AI-made decisions. In other words, Explainable AI helps you "prove the work"—from calculation to decision-making—of an AI. With Explainable AI, you can see what goes into decision-making—and, ultimately, to trust and understand the AI.

How to Explain Your AI Models

Whether you are a data scientist or a business decision maker it is important to uncover how you create trust and understanding of the AI. Some questions you may have:

- AI can make decisions, but can I trust, or more importantly, how do I understand those decisions and interpret them? You need to trust that the AI models are making the right choices because you are ultimately accountable for the decisions.
- How do you “prove the work” of AI? You need to rationalize decisions for auditors and regulators—such as show steps to decision-making, show the transparency behind the models, or provide reason codes.

What you are really asking is “What is my AI thinking?”—or, even more basic, “Can I understand the AI?”

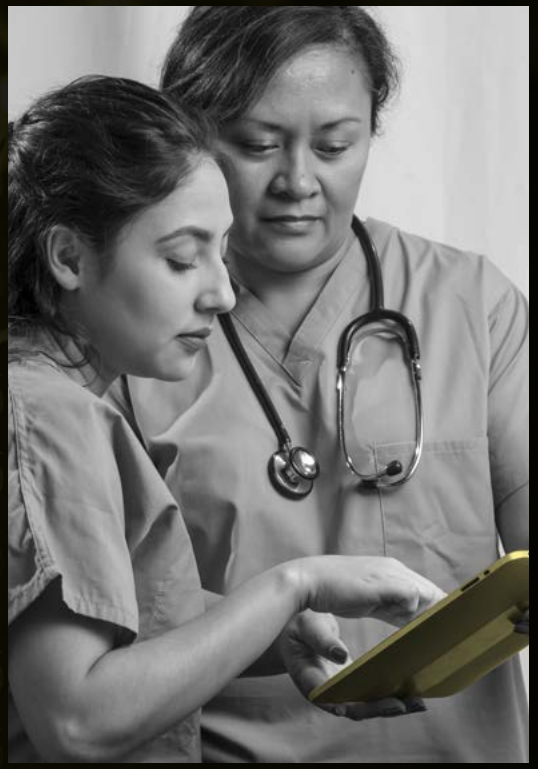
How H2O Driverless AI Automates Explainable AI for Your Business

H2O Driverless AI is an award-winning platform for automatic machine learning that empowers data science teams to scale by dramatically increasing the speed to develop highly accurate predictive models, and then using machine learning interpretability (MLI) to explain and understand the models. H2O Driverless AI provides robust interpretability of machine learning models to explain modeling results.

In the MLI view, H2O Driverless AI employs a host of different techniques and methodologies for interpreting and explaining the results of its models. Four charts are generated automatically including: K-LIME, Shapley, Variable Importance, Decision Tree, Partial Dependence, and Disparate Impact Analysis.

H2O Driverless AI and the machine learning interpretability capability is able to automatically generate the following powerful and transparent outputs:

Reason Codes. Reason codes are required for certain industries including financial services, when providing credit decisions, or for diagnostics in healthcare. Driverless AI can automatically generate these reason codes and show how the key variables influenced the model's prediction—at granular individual level. For example, when declining a new credit card for a specific consumer, the reason codes could show that a recently missed payment, short credit history, and low credit score were the critical factors in the decision.



Assess Fairness. How do your model predictions affect different groups of people? How can the model's predictions across sensitive demographic segments (such as ethnicity, gender, or disability status) be unbiased and fair? Unfairness can materialize in many ways and from many different sources. The process is complex. Disparate impact analysis in Driverless AI can help discuss and handle observational fairness, quickly.

Perform Model Documentation. Driverless AI produces model documentation automatically—a best practice, and a requirement in some industries. AutoDoc, a feature of Driverless AI, includes essential information about machine learning models, such as creation date, creator of the model, intended business purpose, description of the input data set, algorithms tested, model validation steps, and more.

Talk to H2O.ai about Explainable AI

Explainable AI is the next logical step in AI and assists in justifying and even itemizing the steps to decision-making along the way. Explainable AI helps you show how and why the AI landed on the decision it did—and use that documentation for regulators, consumers, internal business leaders, and other members of the team.

H2O Driverless AI is an award-winning platform for automatic machine learning that empowers data science teams to scale by dramatically increasing the speed to develop highly accurate predictive models, and then using MLI and AutoDoc to explain and understand the models.

Whether you are in healthcare, financial services, insurance, pharmaceuticals or another industry, H2O.ai can help you on your AI journey.

Contact H2O to learn more