



Putting AI Reliability to the Test – Deloitte's AI Qualify Offering for Robust AI

Deloitte's AI evaluation tool AI Qualify assesses the robustness of machine learning models. It subjects the model to a series of tests, examining the resilience, reliability, stability and attack vector vulnerability.

The Need: Building stronger ML models

Every day, Machine Learning (ML) solutions are delivering state-of-the-art results across a wide variety of applications, with accuracy beyond the reach of traditional deterministic models. As ML models increasingly govern critical components of our modern world, their reliability becomes an ever greater concern. This is especially true for ML systems that autonomously take decisions (vs merely advise) – from algorithmic credit scoring to autonomous driving. Despite their general effectiveness, ML-based systems are not immune to failure. It is important to understand their inevitable failure modes, so that we may design ML systems to fail in a predictable

and contained fashion, preventing serious harm, protecting against vulnerability to adversaries. We must strive for AI that is as robust and reliable as the traditional systems it enhances or replaces.

Research into adversarial attacks of convolutional neural networks (image recognition) has exposed even sophisticated models to be overly sensitive to a miniscule degree of noise. This opens up opportunities for unscrupulous adversaries who gain access to a model's inner workings to compromise models, to deliberately alter their behavior. Yet all failure modes will not be triggered by sabotage. More commonly, models will slowly lose predictive power

as the operational data on which they are applied becomes less representative of the data on which the models were trained – a fundamental concern of MLOps. There is no cure-all approach to fixing these limitations for all ML models; each model must be individually tested and tuned. One thing is clear: building stronger ML models starts with isolating and understanding their weaknesses. ▶

Our Solution: AI Qualify

Ensuring AI is robust & reliable is a central principle of Deloitte's definition of Trust-worthy AI. The aiStudio tool "AI Qualify" operationalizes the principle by providing a workbench that tests and verifies the behavior of the ML model under investigation, highlighting existing and potential failure modes.

AI Qualify subjects the model to a series of tests, examining the resilience, reliability, stability and attack vector vulnerability. AI Qualify quantifies model performance along each dimension, capturing it in an individual score as well as an overall measure of robustness. It conducts these analyses at several levels of granularity, providing either a quickly digestible overview or the ability to drill down into targeted areas, as required. AI Qualify also tracks robustness vs predictive power over progressive iterations of model development.

AI Qualify assists model developers identify and address model deficiencies, making for better ML models and a stronger showing of compliance to customer, internal governance and regulatory requirements. It accepts any type of classification or regression model, currently focusing on tabular data, with image and text processing in development.

Advantages/Benefits of Deloitte AI Qualify

- Methodical, structured examination of ML model robustness – a central concern for MLOps
- Confidently apply ML models in critical areas with thorough understanding of failure modes
- Ensure the model complies with regulations
- Save valuable time through automated model evaluation

Example Use Cases

- Designing a model to "fail safely" within a given perimeter
- Assessing degree of model generalization/contextual adaptation
- Testing of edge cases or vulnerability to targeted attack
- Tracking model robustness performance over time (development improvements, model drift deterioration)

Contacts

David Thogmartin

Leader aiStudio

Tel: +49 211 8772 2336

dthogmartin@deloitte.de



Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms, and their related entities (collectively, the "Deloitte organization"). DTTL (also referred to as "Deloitte Global") and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/de/UeberUns to learn more.

Deloitte provides industry-leading audit and assurance, tax and legal, consulting, financial advisory, and risk advisory services to nearly 90% of the Fortune Global 500® and thousands of private companies. Legal advisory services in Germany are provided by Deloitte Legal. Our professionals deliver measurable and lasting results that help reinforce public trust in capital markets, enable clients to transform and thrive, and lead the way toward a stronger economy, a more equitable society and a sustainable world. Building on its 175-plus year history, Deloitte spans more than 150 countries and territories. Learn how Deloitte's more than 345,000 people worldwide make an impact that matters at www.deloitte.com/de.

This communication contains general information only, and none of Deloitte GmbH Wirtschaftsprüfungsgesellschaft or Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms or their related entities (collectively, the "Deloitte organization") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.