sigma360

Generative AI and Bias Testing

A Critical Framework for Compliance

Deploying Generative AI in financial compliance requires rigorous bias testing and strong governance. We outline the risks of untested AI, exploring mitigation strategies and responsible adoption to ensure trust and regulatory compliance.

Generative AI (GenAI) solutions promise smarter, more efficient answers to screening at scale, automating and speeding up review processes at unprecedented rates. Yet, while enthusiasm for GenAI keeps growing, **we must also address its potential for algorithmic bias – and find ways to mitigate these implementation challenges**.

The dangers of biased outputs extend far beyond individual analyst workflows. They can lead to widespread consequences such as inconsistent interpretations, inequitable decisions, and even discriminatory practices, bringing potential reputational and regulatory harm to businesses that implement GenAI.

GenAI developers must carry out rigorous, continuous model testing to mitigate these concerns. Without a disciplined testing approach, the very technology designed to enhance control and auditability could inadvertently introduce new, systemic risks.

> Successful GenAI use depends on proactive, rigorous model bias testing.

The Challenge: When Untested AI Undermines Trust

The financial compliance industry operates on a solid foundation of trust and regulation. If businesses deploy GenAI models that have not been through robust bias testing processes, **they could incur severe consequences if any bias is exposed**.

Unequal Scrutiny

Poorly validated models may disproportionately flag entities based on certain characteristics or demographics, leading to discriminatory outcomes. This undermines public trust and increases legal and reputational exposure.

Skewed Risk Profiles

Biases in training data or model design can distort outputs, causing analysts to miss real threats while focusing on low-risk entities. An AI/ML system is generally only as effective as the data used to train it.

Operational Inefficiency

Inaccurate outputs can flood teams with false positives or overlook critical risks, undermining efficiency gains expected from AI adoption.

Regulatory Non-Compliance

Regulatory requirements and expectations for AI ethics are still developing. As AI oversight grows, using inadequately tested models could breach governance standards, inviting reputational damage and legal actions.

Lacking Audit Readiness

Without transparency and traceability, it's difficult to explain or defend automated decisions. Algorithms can generate complex interactions that are difficult to comprehend, leading to unintended discriminatory outcomes.

Sigma360's Approach: Systematic AI Governance

At Sigma360, we recognize that GenAI has great potential to transform risk screening processes from the ground up. We also understand that GenAI must be deployed responsibly. **Teams should maintain control, transparency, and governance while gaining the advantages of AI-powered workflows**.

We have developed and are actively implementing a robust model risk management framework that improves the quality of AI outputs and ensures they can be audited. **Our goal is to reduce bias and maintain model integrity over time**.

Human-in-the-Loop Testing

Human oversight is a crucial part of GenAI testing. Analysts do more than just review AI outputs; their expertise helps AI systems improve continuously. Unlike human investigators, AI systems inherently lack the context and judgment that human operators possess.

At Sigma360, our models undergo a thorough evaluation process across 11 critical dimensions, including coverage, relevance, objectivity and more. This multi-faceted approach goes beyond basic accuracy metrics to ensure systems are operating fairly and reliably across diverse datasets.

> To prevent drift and rectify possible model biases, Sigma360 subject matter experts conduct model performance evaluations and guide refinements over time.

We rigorously test our GenAI models for bias. Between 10,000 and 100,000 prompts are run repeatedly on a global sample of entities representing a wide range of cultures. Outputs are then reviewed for objectivity, specifically checking for negative stereotypes or judgements that are not grounded in the source material. This process directly addresses algorithmic bias by identifying where models may inject subjectivity from their training data.

After this, our experts manually assess potential bias through:

Identifying Loaded Language: We examine outputs for subjective, emotional, or loaded wording to ensure the response remains neutral and evidence-based.

Ensuring Opinions are Sourced: We verify that any opinions are either explicitly requested by our internal prompt or directly extracted from the source material.

Assessment and Refinement: We analyze the lowest-performing prompt examples and refine prompts to reduce bias and improve output consistency.

Ensuring Explainability: Building models that provide explanations for their decisions is critical to identifying and addressing biases.

Our framework acknowledges that while GenAI can process information at unprecedented scale, human expertise remains essential to determining what constitutes bias. Subtle forms of bias may escape standard testing methods, for instance, a model might be accurate overall but still systematically underrepresent or overemphasize certain groups or factors.

At Sigma360, we proactively identify and mitigate these issues before they can affect regulatory compliance.

Comprehensive Documentation

Sigma360 supports model risk management teams by providing exhaustive documentation on:

1 Model Design

We clearly detail the model's **purpose, architectural components, and underlying approach**, ensuring transparency about how our AI models operate and make decisions. Model architecture is guided by a clear definition of desired results, ensuring outcomes can be sufficiently explained.

2 Model Performance

We demonstrate **model effectiveness through robust testing under varied data conditions** across demographic groups and risk scenarios to uncover potential biases. This includes back-testing and forward-testing model accuracy and confirming enhanced performance against conventional approaches.

3 Model Governance

We establish **clear responsibilities, continuous monitoring protocols, and stringent change management processes** to ensure that AI systems remain under organizational control throughout their lifecycle. This includes a well-defined risk appetite that is flexible as AI adoption matures.

This documentation helps teams establish GenAI operations that scale sustainably and adapt to changing business needs while maintaining consistent standards. **Financial institutions bear the responsibility and accountability for the utilization of AI/ML,** encompassing decisions contingent on AI/ML analysis, irrespective of whether these systems are internally developed or externally procured. Overall responsibility lies with the Board and Senior Management.

A Center of Excellence (CoE) or oversight committee can oversee development best practices and establish guidelines for AI system usage. Third-party risk management (TPRM) practices are also critical for managing risks associated with externally sourced AI solutions, covering data security, privacy, and contractual clauses.

Moving Forward With Confidence

Integrating GenAI into compliance brings significant time-saving and automation opportunities. At the same time, teams that leverage GenAI become responsible for properly managing AI models and ensuring ethical use. **Without rigorous bias testing and human oversight, the risks incurred by GenAI could outweigh the benefits**.

Sigma360 keeps AI bias at the forefront of our operations. **We are committed to ensuring our AI models are responsibly managed, and we equip our clients with the transparent documentation required to succeed with GenAI**. We prioritize building trust, fairness, and auditability from the minute our AI systems are implemented. This includes integrating the principles of the three lines of defense model into AI governance frameworks, ensuring clear responsibilities and collaboration for oversight, challenge, and assurance.

To explore a deeper dive into real-world workflows and oversight strategies tailored to your precise risk appetite, we invite you to schedule a consultation or demonstration with our team. We are prepared to help you move forward with confidence.

The *Future* of Financial Crime Compliance Starts with Sigma360

Sigma360 delivers a cohesive, data-centric approach to risk intelligence. Our solutions help firms identify both financial and non-financial risks in real-time at scale.

Compliance teams gain instant access to thousands of integrated data sources, and custom risk scoring, for 1 billion companies and people, making Sigma360 the world's most complete risk analysis platform.

Get Sigma360's AI-curated risk intelligence and taxonomies to boost your team output.

3x
More Data Than Legacy Providers
Covers 600,000+ publications in over 50 different languages, delivering broader, deeper visibility

100B+
Global Risk Signals
Extracted across media, registries, watchlists, and more to identify real risk in real time

260+
Unique Jurisdictions Covered
Global visibility ensures high-risk entities don't slip through jurisdictional blind spots

Sigma360 Solutions

Sanctions & Watchlist Screening
Real-time screening for PEPs, Sanctions, and Adverse Media

Adverse Media Screening
AI-enabled, continous global news screening of entities & people

Perpetual Know Your Customer (KYC)
Ongoing risk mitigation with customizable scoring and alerts

AML Investigations
Global risk intelligence centralized scoring and alerts

Enhanced Due Diligence
Global risk intelligence centralized scoring and alerts

Country Risk Ratings
Expert-designed risk-factor based view into over 260 countries

Counterparty Credit Risk
Comprehensive analysis frameworks, risk intelligence and financial data

White-Glove Onboarding and Optimizations
Receive dedicated, high-touch support from onboarding to ongoing optimizations

Need Customization?
Our team works with you to tailor our flexible features to address your unique business challenges

sigma360