WILMERHALE

WILMER CUTLER PICKERING HALE AND DORR LLP

AI Testing and Monitoring: Lessons from Regulated Industries

Spring Privacy & Security Forum (May 2026)

ATTORNEY ADVERTISING

Panelists & Contact Information

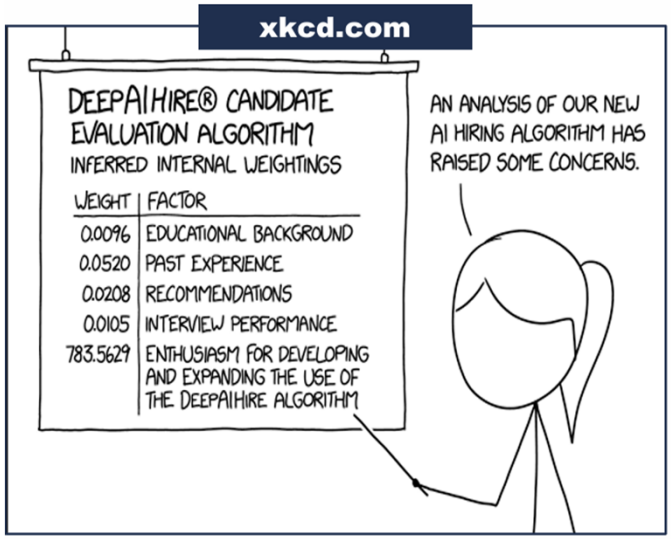
Arianna Evers
Partner
WilmerHale
Arianna.evers@wilmerhale.com
LinkedIn

Nicholas Schmidt
Chief Technology and Innovation Officer
SolasAI
Partner and AI Practice Leader
BLDS, LLC
nschmidt@bldsllc.com
LinkedIn

Chris Meehleib
Associate Director, Responsible AI
UnitedHealth Group
chris.meehleib@uhg.com
LinkedIn

Model Risk

- Potential for adverse consequences from decisions based on incorrect or misused model outputs and reports
- Primarily results from:
  - The model having fundamental errors and producing inaccurate outputs when viewed against the design objective and intended business uses, and
  - The model being used incorrectly or inappropriately
- Model risk increases with greater model complexity, higher uncertainty about inputs and assumptions, broader use, and larger potential impact
- Consider risk from individual models and in the aggregate

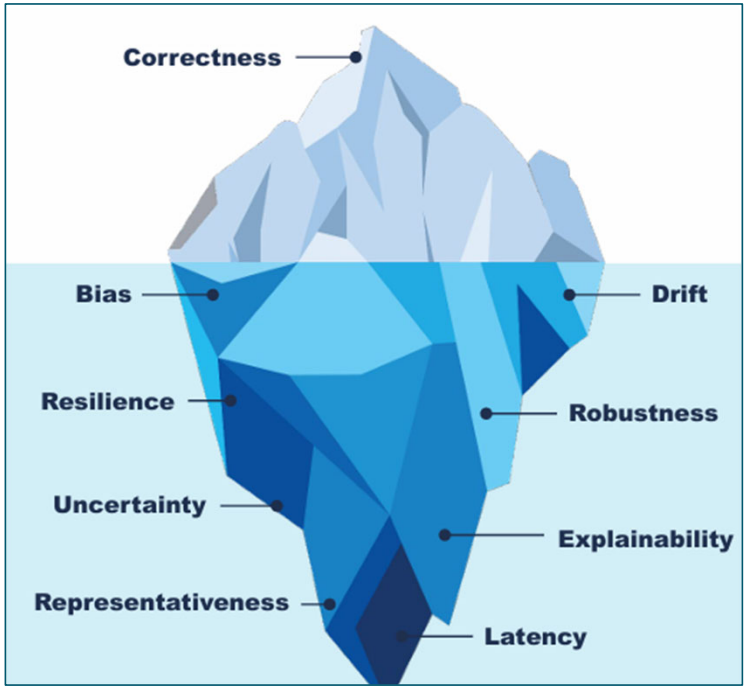


Why test and monitor AI models?

- Legal and regulatory requirements
  - Anti-discrimination laws prohibit disparate treatment and disparate impact even if not specific to AI
  - Other laws and regulations specifically require testing and monitoring
  - Unfair and deceptive trade practices
- Manage model risk and liability
- Support governance and accountability
- Ensure performance and reliability
- Build stakeholder trust

Dimensions of Performance

- Correctness
- Interpretability
- Bias
- Robustness
- Resilience
- Uncertainty
- Drift
- Representativeness
- Latency



Source: Department of Defense, CDAO Test and Evaluation of AI Models (April 2024)

Financial Services

OCC Bulletin 2011-12, "Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management"

- Rescinded by OCC Bulletin 2013-13 (Apr. 17, 2026) and replaced with Model Risk Management: Revised Guidance
- Testing includes:
  - Checking for accuracy, demonstrating that the model is robust and stable, assessing potential limitations, and evaluating model behavior over a range of input values.
  - Assessing the impact of assumptions and identification of situations where the model performs poorly or becomes unreliable
  - Application to actual circumstances under a variety of marketing conditions
  - Appropriate documentation, including test plans (purpose, design, execution), summary results, and detailed analysis
- Ongoing monitoring:
  - Periodically, with a frequency (1) appropriate to the nature of the model, (2) the availability of new data or modeling approaches, and (3) the magnitude of the risk involved
  - Asks whether model components functioning as designed (data inputs, computer code, system integration)
  - Re-testing
  - Analysis of overrides
- Validation of vendor and other third-party products

Financial Services

Monetary Authority of Singapore, Consultation Paper PO17-2025 Consultation Paper on Guidelines on Artificial Intelligence Risk Management (Nov. 2025)

- Evaluation and Testing
  - Relevant and proportionate to the assessed risk materiality
  - Assess performance under a range of plausible conditions
- Document the AI development process to enable reproducibility and auditability, including evaluation measures and performance thresholds, testing approaches, and results
- Post Deployment and Monitoring:
  - Monitoring measures
  - Incident and issue management
  - Roles and responsibilities
  - Documentation
  - Training and Awareness
- High materiality use cases, AI systems, and models may merit regular re-validations by independent parties
- Review aggregate risks across all AI use cases, AI systems, and models

,

Insurance

NAIC Model Bulletin: "Use of Artificial Intelligence By Insurers" (Dec. 2023)

- Provides AI Program guidelines, addressing governance, risk management controls, and internal audit functions
- Oversight of models includes assessments relating to interpretability, repeatability, robustness, regular tuning, reproducibility, traceability, model drift, and auditability of these measurements
- Validating, testing, and retesting as necessary to assess the generalization of the AI system outputs
- For predictive models, a narrative description of the model's intended goals and objectives and how the model is developed and validated to ensure that the AI systems that rely on such models correctly and efficiently predict or implement those goals and objectives
- For third parties, contract terms that provide audit rights and cooperation in regulatory inquiries and investigations
- Regulatory oversight and examination considerations:
  - Processes and procedures related to the management and oversight of Predictive Models, including measurements, standards, or thresholds adopted or used by the Insurer in the development, validation, and oversight of models and AI Systems
  - Documentation related to validation, testing, and auditing, including evaluation of Model Drift to assess the reliability of outputs that influence the decisions made based on Predictive Models.

Insurance

NY Department of Financial Services, Insurance Circular Ltr. No. 7, Use of AI Systems and External Consumer Data and Information Sources in Insurance Underwriting and Pricing (July 11, 2024)

- Frequency of testing: Unfair or unlawful discrimination testing should be administered prior to putting the AI system into production and on a regular cadence thereafter, as well as whenever material updates or changes are made
- Quantitative assessment: Use multiple statistical metrics in evaluating data and model outputs to ensure a comprehensive understanding and assessment (e.g. adverse impact ratio, denials odds ratios, etc.)
- Qualitative assessment: This includes being able to explain, at all times, how the insurer's AIS operates and to articulate a logical relationship between the data/puts and other model variables with an insured or potential insured individual's risk.
- Documentation:
  - Document processes and reasoning behind testing methodologies and analysis for unfair and unlawful discrimination commensurate with the use, materiality, and complexity (and be prepared to make that documentation available to the Department)
  - A description of the process for monitoring data/inputs and AI system usage and performance, including a list of any previous exceptions to policy and reporting
  - A description of testing conducted at least annually to assess the output of AI system models, including any drift that may result from the sue of machine learning or any automated updates
- Vendor agreements: terms that (i) provide audit rights or entitle the insurer to receive audit reports by qualified auditing entities; and (ii) cooperation regarding regulatory inquiries and investigations

Insurance

Colorado Division of Insurance, 3 CCR 702-1 "Governance and Risk Management Framework Requirements for Life Insurers, Private Passenger Automobile Insurers, and Health Benefit Plan Insurers Use of External Consumer Data and Information Sources, Algorithms, and Predictive Models" (Oct. 15, 2025)

- Policies, processes, and procedures, including assigned roles and responsibilities, for the design, development, testing, deployment, use, and ongoing monitoring of data/inputs, as well as algorithms and predictive models that use data/inputs, and processes to ensure that they are documented, tested, and validated.
  - Ensure the data/inputs are credible, relevant, and appropriate for the intended purpose
  - Include an ongoing internal supervision and training program for relevant personnel on the responsible and compliant use of data/puts, and the related algorithms and predictive models.
- Documented description of quantitative testing to detect unfair discrimination, including the methodology, assumptions, results, and steps taken to address unfairly discriminatory outcomes.
- Documented description of ongoing monitoring regarding the performance of algorithms and predictive models that use data/inputs including accounting for model drift.

Standards & Other Resources

NIST AI Risk Management Framework 1.0 and Playbook

- Pre-Deployment Testing
  - Appropriate methods and metrics are identified and applied
  - Evaluate for trustworthy characteristics
  - Track identified AI risks over time
- Ongoing Monitoring
  - Regularly reassess metrics and controls
  - Post-deployment monitoring plans
- Explain, validate, and document the model

NIST AI 800-4, Challenges to the Monitoring of Deployed Systems

Considerations for In-House Counsel

Testing

- Risk-based approach to testing and monitoring
  - Nature of the decisions being made, informed, or supported using the AI system
  - Type and degree of potential harm to consumers
  - Extent to which humans are involved in the final decision-making process
  - Transparency and explainability of outcomes
  - Extent and scope of use or reliance on data, predictive models, or AI systems from third parties
- Focus on performance dimensions relevant to the model and specific use
- More than one test may be appropriate
- Consider independent validation for high-risk models

Considerations for In-House Counsel

Monitoring

- Frequency of ongoing monitoring – annually, periodically, when there is a material update or change to the model, data, or use
- Accountability
- Training
- Incident response

Both

- Documentation
- Legal privilege
- Third-party controls