WBS TRAINING ADVANCED MASTERCLASS SERIES

AI MODEL RISK: Measurement, Reporting and Mitigation

Start date: Tuesday 24th March 2026

4-week flexible global learning

The WBS Advanced Masterclasses run over four weeks in a live online format offering a highly effective learning experience by balancing depth, flexibility, and Faculty engagement.

With a start time of 1pm UK time / 2pm CET / 6:30pm IST / 9:00am EST, enabling students to join live from many locations around the world.

The live format encourages real-time interaction with the instructor and peers, fostering deeper discussions and immediate problem solving. Spacing the sessions weekly also accommodates the busy schedules of banking professionals, ensuring minimal disruption to daily responsibilities while maximising learning retention and practical application. The course ends with a structured assessment multiple-choice test & certificate. Students get to keep the course in their own personal WBS educational portal for 12 months.

Course Overview

START DATE: Tuesday 24th March

DELIVERY MODE:
Live Online (flexible learning over 4 weeks, 1 lecture per week)

GLOBAL START TIME:
Accessible for many participants worldwide.

STRUCTURED ASSESSMENT:
Participants who successfully complete the workshop objectives will be issued a Certificate.

TOTAL LECTURE HOURS: 16 hours

INSTRUCTOR: Alexander Sokol: Head of Quant Research, CompatibL

EMPOWERING GLOBAL TALENT:
Emerging Markets Subsidy

WBS TRAINING PAST DELEGATE:
40% Discount (off the standard fee)

Global Start Times

16 hours: Delivered over 4 weeks, live online, one lecture per week with a 30-minute break, as follows:

Please note: *due to Daylight Saving Time change, the session time on 24th March will begin one hour earlier for UK/Europe, with standardised timing from 31st March onwards.*

Week 1: Tuesday 24th March
1pm UK – live start – 1pm – 5:30pm
2pm CET – live start – 2pm – 6:30pm
6:30pm IST – live start – 6:30pm – 11:00pm
9:00am EDT – live start – 9:00am – 1:30pm

Week 2: Tuesday 31st March
2pm UK – live start – 2pm – 6:30pm
3pm CET – live start – 3pm – 7:30pm
7:30pm IST – live start – 7:30pm – 12:00am
9:00am EDT – live start – 9:00am – 1:30pm

Week 3: Tuesday 7th April
2pm UK – live start – 2pm – 6:30pm
3pm CET – live start – 3pm – 7:30pm
7:30pm IST – live start – 7:30pm – 12:00am
9:00am EDT – live start – 9:00am – 1:30pm

Week 4: Tuesday 14th April
2pm UK – live start – 2pm – 6:30pm
3pm CET – live start – 3pm – 7:30pm
7:30pm IST – live start – 7:30pm – 12:00am
9:00am EDT – live start – 9:00am – 1:30pm

Course Overview

AI model risk is a new discipline, where regulatory requirements and best practices borrow heavily from other types of model risk. The practices adopted from traditional model risk management (MRM) are not always capable of dealing with the unique characteristics and challenges of AI model risk. Effective measurement, reporting and mitigation of AI model risk requires combining novel, AI-specific risk metrics and techniques with traditional risk management practices.

In the first part of the course, Alexander Sokol will present practical and effective techniques for the quantitative measurement and reporting of AI model risk using both well-established and novel metrics. In the second part, Alexander will leverage his award winning research on behavioural psychology of AI to describe practical and effective techniques for mitigating AI model risk and increasing reliability of AI-based workflows.

Each lecture will be accompanied by a hands-on exercise where participants will use an online playground to:

- Implement AI-based workflows based on use cases of practical importance to banking and asset management
- Measure aleatoric and epistemic risk of these workflows
- Mitigate the sources of both types of risk
- Measure the resulting improvements in model risk and reliability metrics

Coding will not be required for workshop participation.

Learning Goals

By the end of the program, participants will:

- Understand quantitative measurement and reporting of AI model risk
- Learn the key metrics for the accuracy and reliability of AI
- Learn how to mitigate AI model risk using advanced techniques from behavioural psychology
- Learn to design reliable AI-based workflows

Technology

The course exercises and hands-on workshop will be available through a browser-based playground. A personal Gemini LLM key will be required to participate. Instructions for getting a Gemini LLM key will be provided. While we do not anticipate any firewall issues, we recommend testing the connectivity to the playground when setup instructions are distributed one week prior to the course start.

Course Schedule And Contents

Week 1: Tuesday 24th March
Introduction to AI model risk

INTRODUCTION, REQUIREMENTS, AND USE CASES

AI model risk as a new discipline
- Rapid adoption of AI in financial services
- Explainability and transparency requirements
- Principles of quantitative measurement and reporting of AI model risk based on rigorous statistical tests

Changes compared to traditional risk management
- Conventional model risk management (MRM) vs. AI model management
  - Periodic validation vs. continuous assurance
  - Traditional backtesting vs. new validation techniques for LLMs
  - The evolving requirements for formal reporting
- Conventional operational risk (OpRisk) vs. AI model risk
  - Challenges of adopting OpRisk metrics for AI model risk
  - Ethical and social implications of using AI in HR and other regulated contexts
- Conventional model stress testing vs. red-team testing
  - Non-adversarial vs. adversarial stress testing approaches for AI model risk
  - Prompt injection and jailbreak testing
  - Robustness to input shifts and edge cases

Reporting requirements for AI model risk
- Regulatory
  - Transparency and explainability documentation
  - Bias and fairness assessment reporting
  - Model inventory and governance oversight
- Internal
  - The importance of quantitative risk metrics based on rigorous statistical tests
  - Continuous monitoring dashboards
  - Drift and performance degradation tracking
  - Use-case specific quantitative risk metrics and KPIs

AI WORKFLOW TYPES

Assistant workflows (multi-turn chat)
- Context management across multiple chats and conversation turns
- Maintaining consistency and coherence
- Handling clarification requests and corrections

Generation workflows (text output)
- Structured vs. free-form text generation
- Template-based and conditional generation
- Quality and style consistency

Comprehension workflows (text input, data output)
- Information extraction and structuring
- Classification and categorisation
- Data validation and quality checks

SELECTED USE CASES

Rating using numerical and category-based scales
- Numeric scales (e.g., 1-5, 1-10, 0-100)
- Category-based scales (e.g., poor/fair/good/excellent)
- Likert scales (degree of agreement)
- Binary classifications (pass/fail, yes/no)

Ranking using pointwise, pairwise, setwise, and listwise approach
- Reliability vs cost trade-offs
- Computational cost and latency considerations
- Bias mitigation requirements for each approach

Complex document analysis using rulebooks
- Security prospectuses
  - Extracting terms and conditions
  - Identifying risk factors and disclosures
  - Assessing the effect of legal caveats
- Regulatory requirements and guidelines
  - Compliance checking against rulebooks
  - Interpretation of ambiguous requirements
- Contracts
  - Key clause identification
  - Obligation and liability extraction
- RFP and RFI questionnaires and responses
  - Requirement matching and scoring
  - Gap analysis and compliance validation

Data entry from free-form text

- Trade confirmations
  - Structured field extraction (dates, amounts, counterparties)
  - Validation against expected formats and ranges
- Free-form emails and chats with trades and market quotes
  - Field recognition by context and position
  - Handling ambiguous semantic structure
  - Handling incomplete information
- Detecting template-generated inputs to improve reliability
  - Pattern recognition for template-generated formats
  - Leveraging template recognition for pre-approval and accuracy improvements

Practical Exercise: Building and testing AI-based workflows

The participants will build and test several AI-based multistep workflows.

Note: No coding required. The exercise will be performed using an online playground.

Week 2: Tuesday 31st March Quantitative Management of AI Model Risk

MEASURING AI MODEL RISK

Statistical analysis of multiple runs

- Sample size and statistical power vs. cost
- Specialized distribution metrics for AI – not just mean and variance
- Confidence intervals and risk reporting

Techniques and challenges of run randomisation

- Temperature and sampling parameter control
- Seed randomisation and reproducibility
- Preamble randomisation to avoid memorisation and as a seed alternative

Systematic vs random errors

- Aleatoric uncertainty (inherent randomness in LLMs)
- Epistemic uncertainty (model capability and knowledge limitations)
- Aleatoric-epistemic decomposition in AI error analysis

Dealing with rare errors and thinking tangents

- Detection methods and reporting for low-frequency errors
- Fast vs. thinking model differences in error patterns
- Monitoring for unexpected reasoning paths

Judge models

- Independent and comparative scoring
- Chain-of-thought prompting for judge models
- Judge model bias detection and mitigation

QUANTITATIVE METRICS BY WORKFLOW TYPE

Measuring rating stability

- Inter-run variance and consistency metrics
- Scale calibration and score distribution analysis
- Central tendency and range of scores

Measuring ranking stability

- Rank correlation metrics (Kendall's tau, Spearman's rho)
- Position bias detection and mitigation
- Agreement rates across multiple runs

Measuring reliability of decision graph navigation for complex document analysis

- Decision path consistency across runs
- Node- and graph-level accuracy metrics
- Error propagation through decision graph nodes

Measuring reliability of data entry for multiple-choice, numerical and other field types

- Accuracy reporting by field type (multiple-choice, numerical, date)
- Detection and mitigation of hallucinated optional fields
- Detection and mitigation of deviations from the output format

Practical Exercise: Measuring AI model risk

The participants will perform quantitative measurement of AI model risk in the workflows they built.

Note: No coding required. The exercise will be performed using an online playground.

Week 3: Tuesday 7th April Mitigation of Psychological Effects and Cognitive Biases in AI Models

PSYCHOLOGICAL EFFECTS

Thinking fast and slow for AI
- System 1 (fast, intuitive) vs System 2 (slow, deliberate) thinking in LLMs
- Failures in cognitive load optimisation
- Switching between System 1 and System 2 in fast models vs. advanced/thinking models
- Chain-of-thought to engage System 2 reasoning

Semantic illusions
- Failures in familiarity detection
- Misleading question structure
- Surface-level vs. deep comprehension testing
- Model susceptibility to deliberate semantic illusions

Framing effects
- Positive vs. negative framing (e.g., rate of success vs rate of failure)
- Influencing risk-averse vs. risk-seeking behaviour in AI
- Stress testing with logically equivalent reformulations

Priming effects
- Influence of unrelated context on responses
- Legal and compliance implications in HR and other regulated contexts
- Mitigation through context randomisation

COGNITIVE BIASES

Confirmation bias, sycophancy, desire to please
- Guessing and meeting user assumptions and expectations at the expense of accuracy
- Seeking information that confirms priors
- Advocating for the perceived user interests

Informational anchoring
- Misinterpretation or over-reliance on initially presented information
- Numeric anchors affecting quantitative outputs
- Testing with varied anchor values

Priming-induced anchoring
- When to expect priming-induced anchoring effects
- Detection and mitigation strategies
- Meeting legal and compliance requirements in HR and other regulated contexts

Central tendency
- Avoiding extreme scores on rating scales in favor of midrange values
- Reduction in ranking stability due to the variable degree of central tendency
- Few-shot and other methods to reduce and stabilize effects of central tendency

Position bias
- Favoring first-presented or last-presented items in multiple evaluation
- Position swap testing methodology and metrics

Practical Exercise: Identifying and mitigating cognitive biases
The participants will identify and mitigate cognitive biases affecting the workflows they built.

Note: No coding required. The exercise will be performed using an online playground.

Week 4: Tuesday 14th April Improving Reliability of AI-Based Workflows

KEY CAUSES OF UNCERTAINTY

Aleatoric vs epistemic uncertainty
- Inherent data randomness vs. model knowledge limitations
- Measurement approaches and mitigation strategies for each type

Hallucinations due to the lack of grounding
- Importance of grounding from external knowledge sources
- Assuming facts learned from common patterns in training data
- Detecting confident but incorrect responses

Psychological effects and cognitive biases
- Impact on reliability and consistency metrics
- Systematic vs. random bias-induced error patterns

Thinking tangents
- Unexpected reasoning paths in complex rules
- Prevention, detection and correction of undesirable thinking tangents

MITIGATION BY PROMPT AND WORKFLOW DESIGN

Challenger models
- Using alternative models for validation
- Cross-model consistency checks
- Identifying model-specific biases and errors

Effective grounding
- Using retrieval-augmented generation (RAG) effectively
- Best practices for using and creating model context protocol (MCP) servers
- Using conventional (non-MCP) knowledge bases
- Citation and web search source tracking

Multistep workflows and decision graphs (rulebooks)
- Breaking complex tasks into manageable steps using rulebooks
- Conditional logic and branching paths
- Error detection and recovery mechanisms for complex rulebooks

Dynamic few-shot
- Selecting relevant examples at runtime
- Similarity ranking-based example retrieval (reverse lookup)
- Identifying and addressing gaps in curated few-shot examples

Corrective few-shot
- Learning from mistakes and failure cases
- Negative examples and counterexamples
- Iterative refinement and improvement
- Mitigation by Monte Carlo

Random sampling across multiple randomized AI model runs as a powerful way to improve AI workflow reliability
- Seed and prefix-based randomisation
- Achieving statistical confidence at a predefined threshold (e.g., 99% confident)

Voting across multiple runs for multiple-choice outputs
- Majority voting and consensus mechanisms
- Weighted voting based on confidence scores
- Handling approximate ties and low-confidence cases

Crowdsourcing across multiple runs for numerical and other continuous-scale outputs
- Effective aggregation techniques in the presence of outliers
- Distribution-agnostic confidence intervals under epistemic and aleatoric uncertainty

Practical Exercise: Using voting and crowdsourcing to improve reliability
The participants will use voting and crowdsourcing to improve reliability of the workflows they built.

Note: No coding required. The exercise will be performed using an online playground.

Course Instructor

Alexander Sokol: Head of Quant Research, CompatibL

Alexander Sokol is the founder and Head of Quant Research at CompatibL, a trading and risk technology company. He is also the co-founder of Numerix, where he served as CTO from 1996 to 2003, and the co-founder of Duality Group, where he served as CTO from 2017 to 2020.

Alexander won the Quant of the Year Award in 2018 together with Leif Andersen and Michael Pykhtin, for their joint work revealing the true scale of the settlement gap risk that remains even in the presence of initial margin. Alexander's other notable research contributions include systemic wrong-way risk (with Michael Pykhtin, Risk Magazine), joint measure models, and the local price of risk (with John Hull and Alan White, Risk Magazine), and mean reversion skew (Risk Books, 2014).

Pricing Structure

STANDARD FEE: £2195.00 + UK VAT

WBS TRAINING PAST DELEGATE: 40% Discount (off the standard fee) – Please submit an enquiry

DISCOUNTS: Early Bird 10% discount available

EMERGING MARKETS DISCOUNT: Please submit an enquiry

GROUP DISCOUNT: If 2 or more people from your institution wish to join the masterclass please contact us.

VAT EXEMPTION: If you are a non-UK or EU resident, you are exempt from UK VAT

Discount Structure

10% Early bird discount

Available until Friday 27th February 2026

40% WBS Training Past Delegate Discount

WBS Training is pleased to offer a 40% special discount for all past delegates at our events and conferences over the past 25 years. Whether you've previously attended our flagship QFC conference, the WQF(A) or other training scenarios, you'll benefit from an exclusive rate designed to reward your ongoing commitment to professional development.

This preferential pricing reflects our appreciation for your dedication to continuous learning and ensures that our expert-level training remains accessible. Don't miss this chance to expand your quant and risk management expertise at a reduced rate, just our way of saying "welcome back" or "keep growing."

Please submit an enquiry.

Empowering Global Talent: Emerging Markets Subsidy

Finance is global. So is talent. That's why WBS Training also offers an Emerging Markets Subsidy to support applicants from developing and underrepresented regions. We believe that potential shouldn't be limited by geography—and this initiative ensures broader access to high-quality financial education.

If you're based in an emerging market and passionate about entering or advancing in the quant finance space, this tailored subsidy is designed for you.

Please submit an enquiry.

AI Model Risk Masterclass – Registration Form

Start Date: Tuesday 24th March 2026

Regular Course Fee

☐ Full Course Fee: £2195.00 + UK VAT

20% VAT IS ONLY CHARGEABLE FOR RESIDENTS IN THE UK AND EU

Early Bird Discount

☐ 10% Discount until 27th February 2026

Discount code

VOLUME DISCOUNT: If 2 or more people from your institution wish to take the masterclass please contact us.

To register, please scan and email the completed booking form to sales@wbstraining.com

DELEGATE DETAILS
NAME:
ORGANISATION:
JOB TITLE:
DEPARTMENT:
ADDRESS:
POSTCODE:
PHONE:
EMAIL:
DATE:
SIGNATURE:

Email: sales@wbstraining.com / Tel: +44 (0) 20 3874 6170