AI Evaluation Methodology Lead

Do you want to define the evaluation methodology and technical requirements that underpin effective AI evaluation and monitoring at one of the leading financial institutions in the Nordics?
Join Danske Bank's newly established AI Model Assessment team in Business Controls & Product Management. As AI Evaluation Methodology Lead, you will hold a 1st Line of Defence role with ownership of the methodologies used to evaluate AI models and AI-enabled solutions, and the technical requirements for platforms supporting evaluation and ongoing monitoring. You will translate business-owned risk and control needs into implementable methods, metrics, data, evidence, traceability and platform capabilities.

Who are we?
The AI Model Assessment team is a business-owned control capability that performs evidence-based assessments of AI models and AI-enabled solutions across the Bank. We work with business owners, developers, engineers and platform teams to understand how solutions operate, define assessment evidence, and establish the evaluation and monitoring capabilities needed to manage model risk.
Our mission is to enable responsible innovation through a consistent, scalable framework for evaluating AI systems, monitoring performance and risk, and demonstrating that appropriate controls operate effectively.

The position is based in Copenhagen, Denmark.

The role
You will own the evaluation methodology and define technical requirements for the platforms, tooling and data capabilities used by the 1st Line to evaluate and monitor AI models and AI-enabled solutions. Your focus is specifying what the capability must deliver—not owning software development or platform operations.
You will work closely with the independent 2nd Line AI model risk function, which sets the overarching risk framework and provides oversight and constructive challenge. You will translate that framework into practical methods, requirements and business-owned controls, provide evidence of implementation, and support the resolution of gaps.

What you will be doing

- Own and maintain evaluation methodologies covering objectives, test approaches, representative datasets, metrics, thresholds, acceptance criteria and evidence standards.
- Define functional and technical requirements for evaluation and monitoring platforms, including data ingestion, test execution, metric calculation, observability, evidence capture, reporting, access controls and auditability.
- Assess how AI architectures, components, dependencies and data flows affect evaluation scope, test design, monitoring coverage and platform requirements.
- Set platform-agnostic requirements across environments such as AWS Bedrock and AgentCore, Databricks, Azure OpenAI and embedded third-party AI platforms.
- Ensure reproducibility and end-to-end traceability across models, prompts, datasets, configurations, components, results, monitoring signals and changes over time.
- Prioritise standardisation and automation opportunities, working with evaluators, platform owners and engineers to turn approved control requirements into practical capabilities.
- Engage with the 2nd Line to interpret requirements, demonstrate operationalisation, respond to challenge and provide evidence supporting independent oversight.

About you

An experienced AI evaluation, model risk, data, architecture or platform professional who can establish rigorous methodologies and convert business risk and control needs into clear, implementable technical requirements.
You have

- Proven experience designing or governing AI/ML evaluation methodologies, including test objectives, datasets, metrics, thresholds, acceptance criteria and evidence standards.
- Experience translating evaluation, monitoring, risk and control needs into requirements for platforms, tooling, data pipelines and reporting.
- Experience with business-owned model risk management and controls, technical assurance, MLOps, LLMOps, AI platforms or AI observability.
- Familiarity with cloud or enterprise AI platforms; experience with AWS Bedrock or AgentCore, Databricks or Mosaic AI, Azure OpenAI, or comparable platforms is advantageous.
- Strong knowledge of representative datasets, experiment configuration, metric design, regression testing, monitoring indicators, traceability and evidence capture.
- Ability to distinguish methodology and requirements ownership from implementation while collaborating effectively with engineers and platform owners. Python or similar analytical and prototyping tools are advantageous, but software development is not the primary focus.
- Excellent communication and collaboration skills across business, assessment, platform, engineering, architecture, data and risk functions.

Experience with AI evaluation or observability technologies, such as Phoenix or comparable tools, is advantageous but not required.

Meet the team

You will join the newly established AI Model Assessment team, combining technical AI evaluation expertise with model risk management across the Bank's AI initiatives. You will connect evaluators with platform, engineering, architecture and data teams, helping build a consistent and scalable business-owned capability while keeping risk ownership and control evidence anchored in the business.

Have we caught your attention?

If you are passionate about robust AI evaluation methodologies and scalable evaluation and monitoring capabilities, we

would love to hear from you.
For questions about the role, please contact Khaled Alizai at khali@danskebank.dk.
Join us and help build the business-owned foundation for effective AI model evaluation, monitoring and risk management at Danske Bank.

Danske Bank	Ansøgningsfrist	Kontaktperson
København K, 1300 København K danskebank.dk	5. november 2026 AI Evaluation Methodology Lead	Khaled Alizai khali@danskebank.dk