JUPITER
CAPITAL MANAGEMENT

EXPERT ADVISORY & CONSULTING SERVICES

Risk Assessments

AI

Risk assessment plays a foundational role in the development of safe, ethical, and trustworthy AI systems by offering a systematic approach to identifying and managing potential harms. As a cornerstone of AI risk management, it enables organizations to proactively address issues such as bias, misuse, and unintended consequences. By evaluating the likelihood of these risks through hazard identification, probability estimation, and the use of both qualitative and quantitative methods, stakeholders can make informed decisions. Tools like risk matrices help prioritize mitigation strategies, while continuous monitoring ensures that risk evaluations remain relevant as AI systems evolve.

IMPACT ANALYSIS

- **Defining Impact Criteria:** Establish metrics such as safety, privacy, financial loss, reputation damage, legal consequences, and ethical considerations.
- **Assessing Severity:** Evaluate how severely the AI system or stakeholders would be affected if the risk occurs, often using qualitative descriptions or quantitative measures.
- **Stakeholder Impact:** Consider the effects on various stakeholders, including users, affected populations, organizations, and society.
- **Impact Levels:** Categorize potential impacts into levels (e.g., low, medium, high) to facilitate prioritization.
- **Scenario Analysis:** Analyze different scenarios of risk occurrence to understand the range of possible impacts.
- **Integration with Likelihood:** Combine impact severity with likelihood estimates to compute overall risk levels, often visualized in risk matrices.

PRIORITIZATION

- **Likelihood Estimation:** Assess the probability of each risk occurring, based on historical data, expert judgment, or scenario analysis.
- **Impact Evaluation:** Determine the severity of consequences if the risk materializes, considering safety, ethical, legal, financial, and reputational effects.
- **Risk Scoring:** Combine likelihood and impact to produce a numerical or categorical risk score, often through methods like risk matrices or weighted formulas.
- **Risk Categories:** Assign risks to predefined levels such as critical, high, medium, or low based on scores, aiding quick identification of priorities.
- **Threshold Setting:** Define specific criteria or cutoff points that differentiate between priority levels, ensuring consistent decision-making.
- **Resource Prioritization:** Allocate mitigation efforts, monitoring, and controls primarily to risks classified as high or critical.
- **Stakeholder Input:** Incorporate insights and concerns from diverse stakeholders to ensure comprehensive prioritization.
- **Dynamic Reassessment:** Continuous updating priorities as new data becomes available or as AI systems and their contexts change.

QUALITATIVE METHODS

- **Expert Judgment:** Gather insights from specialists to evaluate risks based on experience and intuition.
- **Scenario Analysis:** Develop and analyze potential scenarios to understand possible risk outcomes.
- **Risk Categorization:** Classify risks as high, medium, or low based on descriptive criteria.
- **Stakeholder Interviews:** Collect perspectives from various stakeholders to assess risk perceptions and concerns.
- **Checklists and Risk Matrices:** Use predefined lists and visual tools to evaluate and compare risks qualitatively.
- **Descriptive Impact and Likelihood:** Provide narrative descriptions of potential impacts and probabilities instead of numeric estimates.
- **Subjectivity and Bias Management:** Emphasize awareness and mitigation of subjective biases in assessment.

www.jupiter-cap.com **Email: rod.boothe@jupiter-cap.com**

JUPITER
CAPITAL MANAGEMENT

EXPERT ADVISORY & CONSULTING SERVICES

Risk Assessments

AI

Developing responsible AI systems requires a solid risk assessment process that not only promotes safety and ethical use but also minimizes potential societal harm. Central to AI risk management, this process involves identifying possible hazards, estimating their likelihood, and evaluating their impact using structured tools like statistical models, expert input, and scenario analysis. Mapping these risks through a risk matrix and updating assessments through ongoing monitoring allows organizations to prioritize interventions effectively and maintain the integrity of AI systems throughout their lifecycle.

QUANTITATIVE METHODS

- **Data Collection:** Gather numerical data related to risk factors, such as incident rates or failure probabilities.
- **Statistical Modeling:** Apply models like Monte Carlo simulations, Bayesian analysis, or probabilistic risk assessments.
- **Likelihood Quantification:** Assign numerical probabilities to risks based on empirical data or validated models.
- **Impact Quantification:** Measure potential damages or losses in numerical terms (e.g., monetary loss, number of affected individuals).
- **Risk Metrics Calculation:** Compute risk values using formulas like Expected Loss (= Likelihood \times Impact).
- **Sensitivity Analysis:** Assess how changes in variables affect risk levels, useful for identifying key risk drivers.
- **Risk Forecasting:** Use historical data and models to predict future risk levels and trends.

SCENARIO ANALYSIS

- **Scenario Definition:** Craft detailed, plausible situations that could impact the AI system, such as data breaches, malicious manipulation, or unintended consequences.
- **Key Variables Identification:** Pinpoint critical factors influencing the scenario, such as environmental conditions, user behavior, or adversarial actions.
- **Stakeholder Involvement:** Engage relevant stakeholders to ensure scenarios are comprehensive and realistic.
- **Likelihood Estimation:** Assess how probable each scenario is based on available data, expert judgment, or historical trends.
- **Impact Assessment:** Evaluate the potential consequences of each scenario, including safety, ethical, legal, or reputational impacts.
- **Risk Evaluation:** Combine likelihood and impact to determine the risk level associated with each scenario.
- **Mitigation Strategies Development:** Design actions and controls to reduce risks identified in each scenario.
- **Documentation and Communication:** Clearly record scenarios, assumptions, and findings for transparency and stakeholder understanding.
- **Regular Review and Update:** Revisit scenarios periodically to incorporate new insights, emerging threats, and system updates.

STAKEHOLDER INPUT

- **Stakeholder Identification:** Recognize all relevant parties affected by or involved in the AI system, such as users, developers, regulators, and impacted communities.
- **Engagement and Consultation:** Actively involve stakeholders through interviews, surveys, workshops, or focus groups to gather their perspectives on risks and concerns.
- **Risk Perception Collection:** Understand stakeholders' perceptions, fears, and priorities related to AI risks, which may differ from technical assessments.
- **Feedback and Validation:** Incorporate stakeholder insights to validate risk assessments, ensuring they reflect real-world concerns and societal values.
- **Priority Setting:** Use stakeholder input to help prioritize risks, especially those related to ethical, social, or legal considerations.
- **Transparency and Communication:** Clearly communicate risk findings and decision-making processes to stakeholders, fostering trust and accountability.
- **Incorporation into Decision-Making:** Integrate stakeholder perspectives into mitigation strategies, policy development, and governance structures.
- **Continuous Engagement:** Maintain ongoing dialogue to capture evolving concerns and feedback as AI systems develop and deployment contexts change.

www.jupiter-cap.com **Email: rod.boothe@jupiter-cap.com**