Abstract: The increasing deployment of Artificial Intelligence (AI) in high-impact domains such as finance, healthcare, transportation, cybersecurity, and critical infrastructure has intensified the need for risk-aware, trustworthy, and reliable decision-making systems. Traditional machine learning models, while effective for prediction, often struggle with uncertainty, rare events, adversarial conditions, and ethical considerations. This review provides a comprehensive overview of recent advances in risk-aware and trustworthy AI, focusing on approaches that incorporate robustness, interpretability, and governance. It examines key techniques including deep learning-based anomaly detection, uncertainty-aware predictive modeling, generative AI for stress testing, optimization-driven learning pipelines, and explainable AI frameworks. Applications across multiple domains are analyzed to highlight common challenges and design principles. The study also addresses critical issues such as ethics, privacy, and regulatory compliance as integral components of trustworthy AI systems. The paper concludes with future research directions aimed at developing resilient, transparent, and human-centric AI architectures suitable for real-world deployment.

Keywords: Risk-Aware Artificial Intelligence; Trustworthy AI; Anomaly Detection; Generative AI; Explainable AI; Secure Machine Learning; Ethical AI Governance