

CRITÉRIOS PARA A AVALIAÇÃO DOS RISCOS DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

CRITERIA FOR ASSESSING THE RISKS OF ARTIFICIAL INTELLIGENCE SYSTEMS

Adriano Sant'Ana Pedra[1]
Faculdade de Direito de Vitória – Vitória/ES, Brasil
adrianopedra@fdv.br

André Hemerly Paris[2]
Faculdade de Direito de Vitória – Vitória/ES, Brasil
hemerlyparis@gmail.com

RESUMO: O presente estudo investiga os riscos associados à crescente adoção de sistemas de Inteligência Artificial (IA) em áreas críticas, como saúde e segurança pública. Com o avanço tecnológico e a expansão do uso de IA, surge a necessidade de estabelecer critérios rigorosos para avaliar e mitigar os riscos potenciais que esses sistemas podem apresentar. O estudo se concentra na análise dos critérios definidos pelo *European Union Artificial Intelligence Act (EU AI Act*), que visa categorizar e regular o uso de IA com base em seu nível de risco, além de colher insights de outros frameworks (como aqueles propostos pelo *National Institute of Standards and Technology*, Au Yeung, Novelli e Floridi, Slattery, e Stackpole). A pesquisa detalha os diferentes níveis de risco previstos pela legislação europeia, que classifica os sistemas de IA em categorias que variam de risco mínimo a risco inaceitável. Os sistemas de IA com risco mínimo são aqueles que apresentam poucos ou nenhum impacto adverso, enquanto os de risco inaceitável são proibidos devido ao seu potencial de causar danos significativos ou irreparáveis à sociedade ou aos indivíduos. Entre esses extremos, encontram-se os sistemas de risco médio a alto, que exigem medidas adicionais de conformidade, transparência e supervisão para garantir sua implementação segura. Por fim, o estudo sugere que, embora a regulamentação europeia represente um avanço significativo na governança da

[1] Professor do Programa de Pós-Graduação Stricto Sensu - Mestrado e Doutorado - em Direitos e Garantias Fundamentais da Faculdade de Direito de Vitória - FDV. Doutor em Direito Constitucional (PUC/SP). Mestre em Direitos e Garantias Fundamentais (FDV). Mestre em Física Quântica (UFES). Pós-doutorado em Direito (Universidade de Coimbra). Procurador Federal da AGU. Juiz Titular do TRE-ES. Contato: adrianopedra@fdv.br. OrcID: https://orcid.org/0000-0002-8174-9122.

[2] Doutorando em Direitos e Garantias Fundamentais pela Faculdade de Direito de Vitória (FDV). Professor da Pós-graduação de Governança, Gestão de Riscos e Compliance da FDV. Certified Information Privacy Manager (CIPM) e Encarregado de Proteção de Dados Certificado no Brasil (CDPO/BR) pela International Association of Privacy Professionals (IAPP). Certified Compliance and Ethics Professional (CCEP-I) pelo Compliance Certification Board (CCB). Mestre em Direito Processual pela Ufes. Ll.m em Direito Societário pela Fundação Getúlio Vargas. E-mail: hemerlyparis@gmail.com. OrcID: https://orcid.org/0000-0001-6791-0138.

IA, é crucial que essas diretrizes sejam constantemente revisadas e aprimoradas à medida que a tecnologia evolui, de forma a permitir que os benefícios trazidos pela IA não impliquem em prejuízos para os direitos e liberdades de indivíduos.

Palavras-chave: Governança em inteligência artificial; Avaliação de riscos; Transparência algorítmica; Explicabilidade; Privacidade de dados.

REFERÊNCIAS

AU YEUNG, louis. **Guidance for the Development of AI Risk and Impact Assessments**. Disponível em: https://cltc.berkeley.edu/wp-content/uploads/2021/08/AI_Risk_Impact_Assessments.pdf. Acesso em: 20 ago. 2024.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **AI Risk Management Framework**. Disponível em: https://www.nist.gov/itl/ai-risk-management-framework. Acesso em: 20 ago. 2024.

NOVELLI, Claudio; CASOLARI, Federico; ROTOLO, Antonino; TADDEO, Mariarosaria; FLORIDI, Luciano. Taking AI risks seriously: a new assessment model for the AI Act. **AI & Society**, v. 38, n. 3, p. 729-752, 2023. Disponível em: https://link.springer.com/article/10.1007/s00146-023-01723-z. Acesso em: 20 ago. 2024.

SLATTERY, Peter; SAERI, Alexander K.; GRUNDY, Emily A. C.; GRAHAM, Jess; NOETEL, Michael; UUK, Risto; DAO, James; POUR, Soroush; CASPER, Stephen; THOMPSON, Neil. **The AI Risk Repository**: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. MIT FutureTech, Massachusetts Institute of Technology, 2024. Disponível em: https://cdn.prod.website-files.com/669550d38372f33552d2516e/66bc918b580467717e194940_The%20AI%20Risk%20Repository_13_8_2024.pdf. Acesso em: 20 ago. 2024.

STACKPOLE, Beth. **A framework for assessing AI risk**. Disponível em: https://mitsloan.mit.edu/ideas-made-to-matter/a-framework-assessing-ai-risk#:~:text=A%20framework%20based%20on%20a,about%20its%20safety%20and%20bias. Acesso em: 20 ago. 2024.