

**A ROTINIZAÇÃO DO DANO: DESUMANIZAÇÃO E PERDA DA AGÊNCIA
MORAL NA SOCIEDADE DA IA**

**THE ROUTINIZATION OF HARM: DEHUMANIZATION AND LOSS OF
MORAL AGENCY OF IA SOCIETY**

Gustavo Augusto Benedito Costa, Nelson Robson da Silva

Centro Universitário do Sul de Minas , Varginha, MG,

gustavo.costa3@alunos.unis.edu.br

nelson.silva@professor.unis.edu.br

1 INTRODUÇÃO

O avanço da Inteligência Artificial (IA) revolucionou a sociedade e a engenharia, estabelecendo-se como uma força motriz em sistemas decisórios que abrangem desde o diagnóstico médico e a segurança pública até a gestão de crédito e o consumo de informações. No entanto, essa delegação crescente de funções cognitivas e decisórias às máquinas gerou um complexo conjunto de desafios éticos e sociais, levantando questões cruciais sobre responsabilidade, equidade e as consequências não intencionais da automação em larga escala. O foco desta pesquisa, portanto, se delimita na análise das implicações éticas sistêmicas da automação, indo além da mera eficácia técnica para examinar as profundas implicações morais e humanas dessa transformação.

O problema de pesquisa central que norteia este estudo é: De que maneira a rotinização da decisão por sistemas de Inteligência Artificial impulsionada pela busca por eficiência e pela delegação de responsabilidade humana resulta na perda da agência moral e estabelece a indiferença ética como um traço estrutural na sociedade da IA?

Para abordar essa questão, o objetivo geral é analisar de que maneira a rotinização da decisão por sistemas de IA resulta na perda da agência moral e estabelece a Estrutura da Irresponsabilidade Moral na sociedade da IA.

O objetivo geral deste trabalho é analisar de que maneira a rotinização da decisão por sistemas de Inteligência Artificial, impulsionada pela busca por eficiência e pela delegação de responsabilidade humana, resulta na perda da agência moral e estabelece a Estrutura da Irresponsabilidade Moral na sociedade da IA. Para cumprir este propósito, serão estabelecidos três objetivos específicos: primeiramente, contextualizar a “Banalidade do Mal” de Hannah Arendt como lente teórica essencial para explicar como a ausência de julgamento e a adesão cega ao procedimento se manifestam no contexto da automação algorítmica; em segundo lugar, investigar e descrever como a delegação de responsabilidade e a quantificação do indivíduo atuam como mecanismos estruturais que desumanizam e inibem o exercício da agência moral, atingindo tanto os operadores quanto os usuários da tecnologia; e, por fim, demonstrar, por meio de estudos de caso relevantes (como sistemas de justiça ou redes sociais), como a rotinização do dano perpetua desigualdades sociais e consolida a indiferença ética como uma consequência funcional da arquitetura da Inteligência Artificial.

O trabalho se justifica pela urgência em reexaminar o conceito de responsabilidade moral frente à automação crescente e pela necessidade de abordar a crise de legitimidade e capacidade decisória emergente. Em termos acadêmicos, a pesquisa preenche uma lacuna conceitual ao promover um diálogo interdisciplinar entre a Filosofia Política, utilizando o conceito da "Banalidade do Mal" de Hannah Arendt, e a Ética da Tecnologia, oferecendo uma nova lente teórica para compreender como a delegação de responsabilidade e a indiferença ética se tornam traços estruturais da sociedade da IA.

2 REFERENCIAL TEÓRICO

O Referencial Teórico deste trabalho é o alicerce conceitual construído para sustentar a hipótese de que a rotinização do dano é um fenômeno estrutural e previsível na Sociedade da IA, sendo estruturado em três pilares interconectados;

6.1 O Pilar Arendtiano: A Crítica ao Não-Pensamento e o Fundamento da Irresponsabilidade

Neste pilar a Banalidade do Mal é invocada para argumentar que o dano surge da ausência de pensamento e da obediência cega a regras, e não de uma maldade radical, transpondo a crítica arendtiana da burocracia totalitária para o contexto da IA ao entender que a rotinização da decisão algorítmica funciona como a nova burocracia que isenta o agente humano de exercer a faculdade do julgamento.

6.2 Os Mecanismos da Rotinização do Dano: Delegação e Quantificação

O segundo pilar detalha os Mecanismos da Rotinização do Dano: a Quantificação do Indivíduo, conforme descrita por José van Dijck, transforma interações humanas em dados quantificáveis, tornando a decisão algorítmica "matemática" e imune à ponderação ética; esse processo prepara o terreno para a Delegação de Responsabilidade, na qual o conceito de "Moral Crumple Zone" (Zona de Amortecimento Moral) de Madeleine Elish é fundamental ao descrever como o operador humano é forçado a absorver a culpa por falhas do algoritmo, perpetuando o ciclo da irreflexão, e, juntamente com o debate sobre Agentes Morais Artificiais (AMA), se demonstra que a responsabilidade humana sobre o design está sendo ativamente obscurecida.

6.3 O Dano Social como Consequência Estrutural: Crítica e Manifestações

Por fim, o terceiro pilar aborda o Dano Social como Consequência Estrutural, utilizando autores como Safiya Noble e Cathy O'Neil para analisar empiricamente como a Governança Algorítmica, ao ser alimentada por dados históricos enviesados, perpetua vieses sistêmicos e transforma a injustiça do passado em rotina automatizada do futuro, culminando na Síntese Arendtiana-Deleuziana, que conclui que o controle contínuo e a delegação de responsabilidade resultam no desaparecimento do agente moral clássico, produzindo uma subjetividade permanentemente desresponsabilizada.

3 MATERIAL E MÉTODOS

Este trabalho adota uma abordagem metodológica predominantemente qualitativa, de caráter exploratório e analítico-descritivo, com o principal objetivo de construir e validar um esboço teórico que estabeleça uma ponte conceitual entre a teoria filosófica (Hannah Arendt) e a análise crítica dos sistemas sociotécnicos de Inteligência Artificial. O projeto será executado integralmente em ambiente acadêmico, não necessitando de coleta de dados de campo em empresas ou laboratórios, e sua técnica de coleta de dados essencial será a pesquisa bibliográfica e documental, realizada sistematicamente em três frentes principais: primeiro, a análise de obras de Hannah Arendt e filosofia política para fundamentar a crise do julgamento e a perda de agência moral; segundo, a revisão de textos de Teoria Crítica da Tecnologia (focados em Delegação, Quantificação e Dataficação) para descrever os mecanismos da rotinização do dano; e, terceiro, a consulta a relatórios e artigos acadêmicos sobre o dano social em sistemas de IA para contextualizar a relevância empírica do estudo. Por fim, o tratamento dos dados será interpretativo, focado em extrair das fontes os argumentos que conectam a eficiência da rotina algorítmica com a irreflexão ética humana, garantindo a consolidação completa do Referencial Teórico para a análise de caso posterior.

4 RESULTADOS ESPERADOS

Espera-se que este trabalho forneça uma contribuição significativa ao apresentar um traçado teórico robusto que articule a crítica filosófica de Hannah Arendt com os estudos de Ética em IA. O principal produto da pesquisa será a validação da hipótese de que a Rotinização do Dano é o resultado previsível de uma Estrutura de Irresponsabilidade Moral na sociedade da IA. Em termos práticos, a pesquisa deve identificar e mapear os mecanismos pelos quais a delegação de responsabilidade e a quantificação do indivíduo atuam para suprimir a agência moral dos agentes,

demonstrando, através da análise de casos em sistemas de justiça ou redes sociais, como a indiferença ética se torna uma característica funcional e amplificadora de desigualdades. Os impactos esperados incluem fornecer subsídios conceituais para reguladores de tecnologia, incentivando a reintrodução do princípio do julgamento e da reflexão humana no desenvolvimento e fiscalização de sistemas algorítmicos.

REFERÊNCIAS

ARENDDT, Hannah. Eichmann em Jerusalém. Um relato sobre a banalidade do mal. São Paulo: Companhia das Letras, 1999.

ELISH, Madeleine Clare. **Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction** (pre-print). Rochester, NY: Social Science Research Network, 1 mar. 2019. Disponível em: <http://dx.doi.org/10.2139/ssrn.2757236>.

GUAN, H.; **DONG**, L.; **ZHAO**, A. **Ethical risk factors and mechanisms in artificial intelligence decision-making**. Behavioral Sciences, v. 12, n. 9, p. 343, 2022. Disponível em: <https://doi.org/10.3390/bs12090343>.

LAMB, Luís Da Cunha. **Ética em IA e IA ética: prolegômenos e estudo de casos significativos**. Revista USP, n. 141, p. 107–120, 5 jun. 2024. Disponível em: <https://jornal.usp.br/wp-content/uploads/2024/05/8-Luis-Lamb-certo.pdf>.

LO PIANO, Samuele. **Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward**. Humanities and Social Sciences Communications, v. 7, n. 1, p. 9, 17 jun. 2020. Disponível em: <https://doi.org/10.1057/s41599-020-0501-9>.

LOPES, Carolina de Melo Nunes; **MENDES**, Júlia Castro. **Ética e inteligência artificial: desafios e melhores práticas**. Revista da Universidade Federal de Minas Gerais, v. 30, 7 dez. 2023.

MEJIAS, Ulises A.; **COULDRY**, Nick. **Datafication**. Internet Policy Review, v. 8, n. 4, 29 nov. 2019. Disponível em: <https://doi.org/10.14763/2019.4.1428>.

MIRANDA, Fernando Silveira Melo Plentz. O Sentido da Maldade na Obra Eichmann em Jerusalém: um relato sobre a banalidade do mal de Hannah Arendt. Revista Eletrônica Direito, Justiça e Cidadania, São Paulo, v. 4, n. 1 dez., p. 55- 76, 2013.

RIOS, Maurício Cavalcante. A AGÊNCIA MORAL ENTRE A INTELIGÊNCIA ARTIFICIAL E ROBÔS. PÓLEMOS – Revista de Estudantes de Filosofia da Universidade de Brasília, v. 13, n. 30, p. 78–101, 2024. Disponível em: <https://doi.org/10.26512/pl.v13i30.55527>.

SCHUSTER, Nick; LAZAR, Seth. Attention, moral skill, and algorithmic recommendation. Philosophical Studies, v. 182, n. 1, p. 159–184, 1 jan. 2025. Disponível em: <https://link.springer.com/article/10.1007/s11098-023-02083-6>.

VAN DIJCK, José. Artigo Dataficação, dataísmo e dataveillance: Big Data entre o paradigma científico e a ideologia. [S.d.]. Disponível em: <https://cts2017blog.wordpress.com/wp-content/uploads/2017/04/dataficacao-dataismo-dataveillance-van-dijck.pdf>.

ZHANG, Jie; ZHANG, Zong-ming. Ethics and governance of trustworthy medical artificial intelligence. BMC Medical Informatics and Decision Making, v. 23, n. 1, p. 7, 13 jan. 2023. Disponível em: <https://doi.org/10.1186/s12911-023-02103-9>.