

Ética em Inteligência Artificial: Desenvolvimento de um Instrumento Paradidático

Dante Augusto Couto Barone^{1,5}, Rodrigo Mor Malossi¹, Anderson Priebe Ferrugem², Daniel Leal Racheli da Silveira⁴, Akira Borba Colen França⁴, Tatiane Sequerra Stivelman¹

¹Instituto de Informática – UFRGS – Porto Alegre, RS, Brasil

²Computer Engineering Department, Engineering Faculty
Zonguldak Bülent Ecevit University – Zonguldak, Turkey

³Centro de Desenvolvimento Tecnológico – UFPel – Pelotas, RS, Brasil

⁴Departamento de Sistemas e Computação – FURB – Blumenau, SC, Brasil

⁵Programa de Pós-Graduação em Informática na Educação – UFRGS – Porto Alegre, RS, Brasil

barone@inf.ufrgs.br, rodrigo.malossi@ufrgs.br, ferrugem@inf.ufpel.edu.br,
daniel.racheli@ufrgs.br, akirabcf@outlook.com, tatiane.stivelman@inf.ufrgs.br

Abstract. *This paper presents a research instrument for technology students aimed at promoting debate on ethical issues in Artificial Intelligence (AI). The instrument uses Likert-scale and open-ended questions addressing privacy, work, and social consequences of AI. An initial case study reveals participants' concerns about data privacy, impacts on work, and social consequences while acknowledging potential AI benefits. The study concludes by highlighting the importance of interdisciplinary ethical discussions in Computer Science courses.*

Resumo. *Este artigo apresenta um instrumento paradidático destinado a estudantes de tecnologia, visando promover o debate sobre questões éticas em Inteligência Artificial (IA). O instrumento combina questões tipo Likert e dissertativas, abordando três dimensões críticas: privacidade, trabalho e consequências sociais da IA. Um estudo de caso inicial com 70 estudantes revela preocupações significativas sobre privacidade de dados, impactos no trabalho e consequências sociais, enquanto reconhece potenciais benefícios da IA. O trabalho conclui destacando a importância de discussões éticas interdisciplinares nos cursos de Computação.*

1. Introdução

O desenvolvimento e uso ético de sistemas de Inteligência Artificial (IA) têm ganhado destaque internacional. Uma recente revisão sistemática sobre ética em IA generativa destaca como esta tecnologia pode tanto exacerbar quanto reduzir desigualdades socioeconômicas existentes [Hagendorff 2024], ressaltando a necessidade de transparência [Bommasani et al. 2023] e responsabilidade no seu desenvolvimento. Marcos orientadores recentes, como a Recomendação sobre a Ética da Inteligência Artificial da UNESCO [UNESCO 2021] e o EU AI Act [Parliament and of the European Union 2024], enfatizam o respeito aos direitos humanos, à diversidade e à inclusão, estabelecendo requisitos específicos para sistemas de IA de alto risco.

Nesse contexto, este trabalho propõe um instrumento de pesquisa voltado para estudantes de tecnologia, com o objetivo de fomentar o debate sobre questões éticas nas aplicações de IA. O instrumento se fundamenta na teoria de instrumentos paradidáticos, que visa facilitar o processo de ensino-aprendizagem por meio de materiais complementares que estimulam reflexão crítica. As questões objetivas utilizam a escala tipo Likert [Antoniali et al. 2015], escolhida por sua eficácia comprovada

na mensuração de atitudes e opiniões. O instrumento está estruturado em três dimensões fundamentais relacionadas à ética em IA: privacidade, relações de trabalho e consequências sociais.

2. Privacidade

O uso crescente de dados pessoais no desenvolvimento de sistemas de IA tem gerado preocupações significativas sobre privacidade e proteção de dados. Estudos recentes demonstram que é possível extrair dados sensíveis de modelos de linguagem, como o ChatGPT, sem conhecimento prévio do conjunto de dados de treinamento [Nasr et al. 2023]. Hagendorff [Hagendorff 2024] identificou a privacidade como uma das principais preocupações éticas relacionadas à IA generativa, destacando riscos de vazamento de dados e necessidade de maior transparência. Sistemas de IA podem violar direitos fundamentais previstos na Declaração Universal dos Direitos Humanos [Cassino et al. 2023], especialmente quanto à privacidade (Artigo 12) e proteção contra discriminação (Artigo 7).

No contexto brasileiro, a Lei Geral de Proteção de Dados Pessoais (LGPD) [da República do Brasil 2018], embora não aborde diretamente a IA, estabelece princípios aplicáveis como finalidade específica, adequação aos objetivos informados e transparência. Preocupações adicionais incluem viés social em modelos de geração de imagens [Wang et al. 2024] e vulnerabilidade a ataques adversariais [Zou et al. 2023], que podem manipular modelos para gerar conteúdo indesejado. Pesquisas revelaram como algoritmos de análise comportamental podem inadvertidamente expor dados sensíveis dos usuários, incluindo orientação sexual e afiliação política [Kosinski et al. 2016].

3. Relações de Trabalho

A automação de tarefas complexas por sistemas de IA tem gerado debates significativos sobre o futuro do trabalho. Estudos do Fórum Econômico Mundial preveem que até 2025, 85 milhões de empregos podem ser deslocados pela automação, enquanto 97 milhões de novos papéis podem emergir [World Economic Forum 2020]. Este cenário demanda uma urgente remodelação das políticas educacionais e trabalhistas, especialmente em países em desenvolvimento. Uma análise abrangente da literatura [Hagendorff 2024] destaca a importância de considerar não apenas a substituição de funções, mas também os benefícios potenciais da complementaridade entre humanos e máquinas.

No Brasil, a ausência de legislação específica sobre automação, prevista no artigo 7º, XXVII, da Constituição Federal, dificulta a proteção dos trabalhadores. Esta lacuna regulatória é especialmente crítica considerando as características do mercado de trabalho latino-americano, marcado por alta informalidade e desigualdade social. Pesquisas recentes identificaram três categorias principais de impacto da IA no trabalho: substituição por automação completa, aumento como ferramenta para melhorar o desempenho humano e transformação através da criação de novas profissões.

4. Consequências Sociais

O impacto da IA no tecido social vai além de questões de privacidade e trabalho, afetando fundamentalmente como processamos informações e tomamos decisões coletivas. Um exemplo crítico é o avanço das tecnologias de deepfakes, que criam desafios significativos para a verificação de evidências e a confiabilidade da informação. Hagendorff [Hagendorff 2024] destaca que os impactos sociais da IA generativa incluem preocupações sobre desinformação, manipulação de mídia e a necessidade de maior regulamentação, aspectos que se alinham com as preocupações identificadas em nossa pesquisa. No contexto eleitoral brasileiro, a cooperação técnica entre Anatel e TSE [ANATEL and TSE 2023] representa uma resposta institucional ao desafio da desinformação potencializada por IA, visando criar mecanismos ágeis para combater a disseminação de conteúdo falso.

A questão da responsabilização por decisões de IA também emerge como ponto crítico. Casos recentes, como acidentes com veículos autônomos, discriminação em sistemas de análise de crédito [Wang et al. 2024], e erros em sistemas médicos automatizados [Omiye et al. 2023], evidenciam a

complexidade da cadeia de responsabilidade. Essas situações demonstram a necessidade de marcos regulatórios que estabeleçam claramente a cadeia de responsabilidade, desde desenvolvedores até usuários finais, garantindo uma abordagem ética e segura para o uso da IA.

5. Metodologia

O instrumento paradigmático para discussões éticas sobre IA foi desenvolvido na disciplina Ethics in Artificial Intelligence da Universidade Federal do Rio Grande do Sul, envolvendo estudantes de graduação em Ciência da Computação e pós-graduação em Informática na Educação, o que permitiu uma abordagem multidisciplinar. A pesquisa foi conduzida com 70 estudantes, majoritariamente de Ciência da Computação e Engenharia de Computação. O instrumento é caracterizado como paradigmático por complementar o currículo formal, estimular reflexão crítica e contextualizar situações reais, utilizando uma combinação de questões tipo Likert (escala 1-5) e dissertativas. A escala Likert foi escolhida por sua eficácia comprovada em mensurar atitudes, enquanto as questões foram estruturadas seguindo critérios de validação, garantindo uma sequência lógica e inter-relação entre os itens. Além disso, as questões dissertativas foram incluídas para explorar em profundidade as percepções dos estudantes sobre temas como privacidade, impacto no trabalho e consequências sociais da IA.

6. Resultados e Discussão

A análise dos dados, fundamentada na escala Likert [Joshi et al. 2015] e respostas dissertativas (Tabela 1), revelou preocupações significativas em três eixos principais: privacidade, relações de trabalho e consequências sociais. Em privacidade, os participantes demonstraram forte demanda por transparência no uso de dados e criticaram as atuais medidas de segurança; já no contexto trabalhista, identificou-se uma opinião polarizada sobre a legitimidade da substituição por IA, com ênfase na necessidade de complementaridade entre tecnologia e trabalho humano, e preocupações sobre potencial aumento da desigualdade social; nas consequências sociais, emergiu forte preocupação com desinformação, manipulação tecnológica e lacunas regulatórias, evidenciando a complexidade ética do avanço da inteligência artificial.

Categoria	Questão	Respostas				
		5	4	3	2	1
Privacidade	Empresas devem ser transparentes sobre uso de dados	58	10	2	0	0
	Dados devem ser anonimizados	51	8	10	0	1
	Medidas atuais de segurança são suficientes	1	1	17	20	31
Trabalho	Importante treinar trabalhadores para novas tecnologias de IA	36	22	9	3	0
	Substituição de pessoas por IA é legítima	18	15	14	12	11
Consequências Sociais	Deepfakes para desinformação é preocupante	57	6	3	4	0
	IA deve ter limites claros em decisões pessoais	57	9	2	1	1

Tabela 1. Exemplo de perguntas e respostas sobre Percepções de IA separados por categorias

7. Conclusão e Trabalhos Futuros

Este estudo apresentou e validou um instrumento paradigmático para fomentar discussões éticas sobre Inteligência Artificial (IA) entre estudantes de tecnologia. Os resultados mostraram a eficácia do instrumento tanto como ferramenta de pesquisa quanto como recurso pedagógico, destacando três contribuições principais: o desenvolvimento de um instrumento validado para avaliar percepções éticas sobre IA, a criação de material paradigmático para discussões de ética em IA, e o mapeamento

inicial das percepções de estudantes brasileiros sobre questões éticas relacionadas à IA. Os resultados revelaram uma consciência crítica sobre as implicações éticas da IA, com preocupações específicas em dimensões como privacidade, trabalho e consequências sociais.

As descobertas do estudo estão alinhadas com diretrizes internacionais, como o AI Act europeu e as recomendações da UNESCO, sugerindo a necessidade de fortalecer o arcabouço regulatório, integrar discussões éticas nos currículos de computação e desenvolver mecanismos de governança para IA. No entanto, o estudo apresenta limitações, como a amostra restrita a três instituições de formas não balanceada, o foco em estudantes de computação e o período limitado de coleta de dados. Para trabalhos futuros, sugere-se a expansão do estudo para outras instituições, o desenvolvimento de módulos educacionais baseados nos resultados, estudos longitudinais para acompanhar a evolução das percepções e a adaptação do instrumento para outros públicos, permitindo uma compreensão mais ampla e aprofundada das percepções sobre ética em IA em diferentes contextos.

Referências

- ANATEL and TSE (2023). Acordo de cooperação técnica nº9/2023. Tribunal Superior Eleitoral.
- Antoniali, F., Antoniali, L. M., and Antoniali, R. E. (2015). Usos e abusos da escala likert: Estudo bibliométrico nos anais do enanpad de 2010 a 2015.
- Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D., and Liang, P. (2023). The foundation model transparency index. *arXiv*.
- Cassino, J., Alvero, A. J., and Silveira, A. C. (2023). Undefined title.
- da República do Brasil, P. (2018). Lei geral de proteção de dados pessoais (LGPD).
- Hagendorff, T. (2024). Mapping the ethics of generative ai: A comprehensive scoping review. *arXiv preprint arXiv:2402.08323*.
- Joshi, A., Kale, S., Chandel, S., and Pal, D. K. (2015). Likert scale: Explored and explained. *Current Journal of Applied Science and Technology*, 7(4):396–403.
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., and Stillwell, D. (2016). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 71(6):543–556.
- Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., and Lee, K. (2023). Scalable extraction of training data from (production) language models. *arXiv*.
- Omiye, J. A., Lester, J. C., Spichak, S., Rotemberg, V., and Daneshjou, R. (2023). Large language models propagate race-based medicine. *npj Digital Medicine*, 6(1):1–4.
- Parliament, E. and of the European Union, C. (2024). Regulamento da inteligência artificial (ue) 2024/1689.
- UNESCO (2021). Recomendação sobre a Ética da inteligência artificial. Recomendação, Organização das Nações Unidas para a Educação, a Ciência e a Cultura, Paris.
- Wang, W., Bai, H., Huang, J., Wan, Y., Yuan, Y., Qiu, H., Peng, N., and Lyu, M. R. (2024). New job, new gender? measuring the social bias in image generation models. *arXiv*.
- World Economic Forum (2020). The future of jobs report 2020.
- Zou, A., Wang, Z., Carlini, N., Nasr, M., Kolter, J. Z., and Fredrikson, M. (2023). Universal and transferable adversarial attacks on aligned language models. *arXiv*.