

# O uso de redes neurais artificiais para medir a participação de mulheres em pesquisas científicas sobre Inteligência Artificial

Afonso Cesar Lelis Brandão  
Programa de Pós-Graduação em  
Engenharia Elétrica e Computação  
Universidade Presbiteriana Mackenzie  
Email: [afonsolelis@gmail.com](mailto:afonsolelis@gmail.com)

Jairo Filho Sousa de Almeida Ferreira  
Programa de Pós-Graduação em  
Engenharia Elétrica e Computação  
Universidade Presbiteriana Mackenzie  
Email: [dealmeida.jairo@gmail.com](mailto:dealmeida.jairo@gmail.com)

**Abstract**—Trabalho feito para a disciplina de redes neurais artificiais do programa de Pós-Graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie. Para realização deste trabalho, foi selecionado como base o artigo "What's in a name? - gender classification of names with character based machine learning models", o artigo foi publicado na revista Data Mining and Knowledge Discovery (ISSN 1573-756X), classificada pela Capes como Qualis A1. DOI - <https://doi.org/10.1007/s10618-021-00748-6>

A identificação de informações bibliométricas é importante, pois elas ajudam a identificar nichos de pesquisa e tendências em determinadas áreas do conhecimento. Além disso, as informações censitárias são relevantes, pois possibilitam a identificação de possíveis vieses na pesquisa, fornecendo uma base sólida para análise e interpretação dos resultados obtidos. É fundamental destacar a importância dessas informações para a produção de conhecimento de qualidade e para o avanço da ciência em diversas áreas [4].

## 1. Introdução

### 1.1. Contextualização

A Inteligência Artificial (IA) tem se destacado como uma tecnologia amplamente utilizada em diversas áreas, desde a saúde até a indústria, proporcionando benefícios significativos. Segundo os indicadores do Google Trends, as buscas relacionadas à IA aumentaram expressivos 269% entre março de 2022 e março de 2023. Dentre as tecnologias acessíveis e amplamente aplicadas na atualidade, destaca-se o ChatGPT da OpenAI, um modelo de linguagem natural utilizado em diversas áreas, tais como atendimento ao cliente, assistentes virtuais e análise de sentimentos [1, 2].

A Inteligência Artificial tem se tornado uma área de grande interesse na pesquisa científica acadêmica, como evidenciado pelo aumento significativo no número de artigos publicados sobre o tema ao longo dos anos. Em 2000, apenas 33 artigos foram publicados na Web of Science, enquanto em 2022 esse número saltou para 4.885, representando um impressionante aumento de 14.803% no período [2, 3].

Os Estados Unidos lideram a produção de publicações científicas sobre Inteligência Artificial, sendo responsáveis por 23,43% dos artigos publicados. A China ocupa o segundo lugar com 20,59%, enquanto a Inglaterra aparece em terceiro lugar com 7,68%. A grande maioria desses artigos foram publicados em língua inglesa, totalizando 98%. Indicadores de extraídos da base Web of Science, permitem a condução de diferentes análises bibliométricas, porém, nenhum dos relatórios emitidos pela plataforma, trazem informações censitárias dos autores [3].

TABLE 1. TABELA DE PUBLICAÇÕES POR AUTOR(A) E GÊNERO

Autor(a)	Publicações	Gênero
hline Liu Y	61	Masculino
Zhang Y	60	Masculino
Li Y	56	Informação indisponível
Li J	53	Masculino
Zhang J	45	Feminino
Wang Y	44	Masculino
Park J	40	Masculino
Kim J	39	Feminino
Wang J	39	Informação indisponível
Kumar A	37	Masculino

A ascensão da Inteligência Artificial na pesquisa científica é um tema que vem recebendo grande destaque. No entanto, outro aspecto importante que merece atenção é a representatividade de gênero dos autores envolvidos nessas pesquisas. Ao analisar os 10 autores com maior quantidade de trabalhos publicados, constatou-se que 60% dos autores são do sexo masculino e 20% são do sexo feminino, outros 20% não foram identificados, evidenciando a baixa participação das mulheres na área [3].

Infelizmente, a sub-representação das mulheres não é uma realidade exclusiva da Inteligência Artificial e do aprendizado de máquina, mas é um problema que permeia várias áreas da ciência. No entanto, é importante ressaltar que a inclusão de mulheres na área de IA não só é fundamental para promover a igualdade de gênero, mas também para melhorar a qualidade e diversidade das soluções em IA. Ao contar com a perspectiva feminina, é possível obter soluções mais criativas e inclusivas, que atendam às necessidades de toda a sociedade [5, 6].

A sub-representação das mulheres em áreas STEM é um

fenômeno global. De acordo com a UNESCO (2021), apenas 35% dos estudantes matriculados em cursos de ciência, tecnologia, engenharia e matemática são mulheres em todo o mundo. Esse desequilíbrio pode ser observado em todas as etapas da carreira, desde a educação até a liderança em empresas e instituições de pesquisa [8].

Vários estudos têm mostrado que a diversidade de gênero nas equipes de pesquisa e desenvolvimento em IA pode melhorar a qualidade das soluções em IA e aumentar a inovação. Segundo um relatório da McKinsey Company (2020), as empresas com maior diversidade de gênero em suas equipes são mais propensas a superar seus concorrentes em termos de rentabilidade e desempenho. Além disso, a diversidade de gênero nas equipes de IA pode ajudar a evitar vieses e discriminação em sistemas automatizados.

Diante disso, a inclusão das mulheres em áreas de IA e aprendizado de máquina é fundamental para promover a igualdade de gênero e melhorar a qualidade e diversidade das soluções em IA.

## 1.2. Objetivos

- Desenvolver um modelo de Rede Neural Artificial capaz de classificar o gênero (masculino ou feminino) com base no primeiro nome da pessoa;
- Construir uma base de dados com nome de autores de trabalhos acadêmicos sobre inteligência artificial;
- Identificar o gênero dos autores dos artigos científicos sobre inteligência artificial;
- Realizar uma análise descritiva para avaliar a participação feminina na autoria de artigos científicos sobre inteligência artificial.

## 1.3. Objetivos

## 2. Materiais e métodos

Para identificar o gênero dos autores com base em seus nomes, vamos utilizar o modelo de Rede Neural Artificial proposto por Hu et al. (2021), que é baseado em caracteres e foi desenvolvido para classificar o gênero a partir do nome. Esse modelo é uma rede neural artificial de aprendizagem supervisionada por classificação, ou seja, ele é capaz de classificar dados em diferentes categorias ou classes [9]. Para treinar a rede, serão utilizados dados da Social Security Administration (SSA), que disponibiliza informações sobre a quantidade de pessoas registradas com determinado nome, separadas por gênero (tabela 2).

O processo de treinamento da rede neural de aprendizagem supervisionada por classificação começa com a definição da arquitetura da rede, que inclui o número de camadas, o número de neurônios em cada camada e a função de ativação utilizada. Em seguida, o modelo é exposto a um conjunto de dados de entrada rotulados, juntamente com as classes correspondentes. Durante o treinamento, a rede neural ajusta seus pesos e conexões de modo que seja capaz de classificar corretamente novos dados de entrada.

Após o treinamento do modelo, a rede neural será utilizada para classificar o gênero dos autores em uma base de dados bibliométrica composta por 15.196 artigos completos, criada a partir dos dados extraídos da Web of Science. Para essa tarefa, será utilizada a string de busca TI=(“artificial intelligence”). Levando em conta que 81,83% dessa amostra (12.463 artigos) foi publicada nos últimos cinco anos (2018 a 2022), serão excluídos os trabalhos publicados fora desse intervalo. Com o modelo treinado, os atributos da base de dados utilizada no treinamento permitirão que a rede neural ajuste seus pesos de acordo com as tendências de nomes masculinos e femininos, possibilitando a classificação correta do gênero dos autores. A rede neural de aprendizagem supervisionada por classificação é amplamente utilizada em diferentes áreas, como reconhecimento de imagem, processamento de linguagem natural e análise de dados, e pode fornecer resultados precisos e confiáveis quando adequadamente treinado e validado.

Male name	Count	Female name	Count
James	5.2M	Mary	4.1M
John	5.1M	Elizabeth	1.6M
Robert	4.2M	Patricia	1.6M
Michael	4.4M	Jennifer	1.5M
William	4.1M	Linda	1.5M

TABLE 2. TOP MALE AND FEMALE US BABY NAMES IN SSA DATA.

## 3. Resultados Esperados

Espera-se que a aplicação do modelo de Rede Neural Artificial proposto por Hu et. al (2021) possa identificar a relevância do gênero feminino nas publicações de redes neurais em 2022, com alta acurácia na classificação do gênero de um nome presente em um artigo científico. Os resultados obtidos poderão contribuir para a promoção da diversidade e inclusão na área de redes neurais, fornecendo informações úteis para o desenvolvimento de políticas e ações que incentivem a participação e a representatividade das mulheres. Este estudo pode ainda fornecer novas abordagens para a solução de problemas de classificação de gênero e para a promoção da igualdade de gênero na área de redes neurais.

## References

- [1] Google. (2023). Google Trends. <https://trends.google.com/trends/explore>
- [2] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
- [3] Clarivate Analytics. (2023). Web of Science Core Collection. Recuperado de <https://www.webofknowledge.com>
- [4] Van Raan, A. F. J. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung - Theorie und Praxis*, 12(1), 20-29.
- [5] Stoet, G., Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological science*, 29(4), 581-593.

- [6] Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter?. *Gender and Education*, 17(4), 369-386.
- [7] UNESCO. (2021). *Cracking the code: Girls' and women's education in STEM*. Paris: UNESCO.
- [8] Hunt, V., Prince, S., Dixon-Fyle, S., Yee, L. (2020). *Delivering through diversity*. McKinsey Company.
- [9] Hu, Y., Hu, C., Tran, T., Kasturi, T., Joseph, E., Gillingham, M. (2021). What's in a name? - gender classification of names with character based machine learning models. In *Proceedings of the 2021 IEEE International Conference on Big Data (Big Data)* (pp. 4164-4169). IEEE.
- [10] H. Kabir, M., Ahmad, F., Hasan, M. A. M., Shin, J. (2020). Gender Recognition of Bangla Names Using Deep Learning Approaches. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 12(2-7), 79-83.
- [11] UNESCO. (2021). Fatores que afetam a participação feminina em disciplinas STEM: Uma revisão de literatura. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000376752>
- [12] McKinsey Company. (2020). Diversity wins: How inclusion matters. Retrieved from <https://www.mckinsey.com/business-functions/organization/our-insights/diversity-wins-how-inclusion-matters>.