

Ernane Rosa Martins
(Org.)

INTELIGÊNCIA ARTIFICIAL APLICADA

SOLUÇÕES PARA UM MUNDO AUTOMATIZADO



científica digital



EDITORA CIENTÍFICA DIGITAL LTDA

Guarujá - São Paulo - Brasil
www.cientificadigital.org - contato@cientificadigital.org

Diagramação e Arte Edição © 2025 Editora Científica Digital
Equipe Editorial Texto © 2025 Os Autores
Imagens da Capa 1ª Edição - 2025
Adobe Stock - 2025 Acesso Livre - Open Access

© COPYRIGHT - TODOS OS DIREITOS RESERVADOS. A editora detém os direitos autorais sobre a edição e o projeto gráfico, enquanto os autores mantêm os direitos autorais de seus respectivos textos. Esta obra está licenciada sob a Licença Creative Commons Atribuição 4.0 Internacional, permitindo o download e compartilhamento integral ou parcial, desde que a fonte seja devidamente citada e os créditos atribuídos aos autores. É obrigatório que a obra permaneça em formato de Acesso Livre (Open Access), sem qualquer alteração. A catalogação em plataformas de acesso restrito ou com fins comerciais é estritamente proibida.



Dados Internacionais de Catalogação na Publicação (CIP)

161 Inteligência artificial aplicada: soluções para um mundo automatizado / Organização de Ernane Rosa Martins. – Guarujá-SP: Científica Digital, 2025.

Formato: PDF
Requisitos de sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Inclui Bibliografia
ISBN 978-65-5360-891-7
DOI 10.37885/978-65-5360-891-7

1. Inteligência artificial. I. Martins, Ernane Rosa (Organizador). II. Título.

CDD 006.2

Elaborado por Janaína Ramos – CRB-8/9166

Índice para catálogo sistemático:

I. Inteligência artificial

E-BOOK

ACESSO LIVRE ON LINE - IMPRESSÃO PROIBIDA

2025

Ernane Rosa Martins
(Org.)

Inteligência Artificial Aplicada: Soluções para um Mundo Automatizado

1ª EDIÇÃO



científica digital

2025 - GUARUJÁ - SP

CONSELHO EDITORIAL

Prof. Dr. André Cutrim Carvalho
Prof. Dr. Antônio Marcos Mota Miranda
Prof^a. Ma. Auristela Correa Castro
Prof. Dr. Carlos Alberto Martins Cordeiro
Prof. Dr. Carlos Alexandre Oelke
Prof^a. Dra. Caroline Nóbrega de Almeida
Prof^a. Dra. Clara Mockdece Neves
Prof^a. Dra. Claudia Maria Rinhel-Silva
Prof^a. Dra. Clecia Simone Gonçalves Rosa Pacheco
Prof. Dr. Cristiano Marins
Prof^a. Dra. Cristina Berger Fadel
Prof. Dr. Daniel Luciano Gevehr
Prof. Dr. Diogo da Silva Cardoso
Prof. Dr. Ernane Rosa Martins
Prof. Dr. Everaldo dos Santos Mendes
Prof. Dr. Fabricio Gomes Gonçalves
Prof^a. Dra. Fernanda Rezende
Prof. Dr. Flávio Aparecido de Almeida
Prof^a. Dra. Francine Náthalie Ferraresi Queluz
Prof^a. Dra. Geuciane Felipe Guerim Fernandes
Prof. Dr. Humberto Costa

Prof. Dr. Joachin Melo Azevedo Neto
Prof. Dr. Jónata Ferreira de Moura
Prof. Dr. José Aderval Aragão
Prof. Me. Julianno Pizzano Ayoub
Prof. Dr. Leonardo Augusto Couto Finelli
Prof. Dr. Luiz Gonzaga Lapa Junior
Prof. Me. Marcelo da Fonseca Ferreira da Silva
Prof^a. Dra. Maria Cristina Zago
Prof^a. Dra. Maria Otília Zangão
Prof. Dr. Mário Henrique Gomes
Prof. Dr. Marcus Fernando da Silva Praxedes
Prof. Dr. Nelson J. Almeida
Prof. Dr. Pedro Afonso Cortez
Prof. Dr. Reinaldo Pacheco dos Santos
Prof. Dr. Rogério de Melo Grillo
Prof^a. Dra. Rosenery Pimentel Nascimento
Prof. Dr. Rossano Sartori Dal Molin
Prof. Me. Silvio Almeida Junior
Prof^a. Dra. Thays Zigante Furlan Ribeiro
Prof. Dr. Wescley Viana Evangelista
Prof. Dr. Willian Carboni Viana
Prof. Dr. Willian Douglas Guilherme

Acesse a lista completa dos Membros do Conselho Editorial em www.editoracientifica.com.br/conselho

Parecer e revisão por pares

Os textos que compõem esta obra passaram por avaliação do Conselho Editorial e revisão por pares externos (*Peer Review*), recebendo a devida recomendação para publicação.

Nota: Esta obra é fruto de um processo colaborativo, configurando-se como uma coletânea na qual os direitos autorais permanecem resguardados para os respectivos autores. Alguns capítulos podem ter origem em trabalhos anteriormente apresentados em eventos acadêmicos; no entanto, os autores foram orientados a adotar o devido rigor na prevenção do autoplágio. A responsabilidade pelo conteúdo de cada capítulo, assim como pela originalidade e integridade das informações publicadas, é inteiramente dos respectivos autores e autoras. O conteúdo da obra não reflete, necessariamente, a opinião da editora, dos organizadores ou dos membros do conselho editorial.

APRESENTAÇÃO

Esta obra é fruto de um esforço colaborativo que reuniu professores, estudantes e pesquisadores cujo envolvimento enriqueceu significativamente as discussões neste espaço formativo. Além disso, resulta de iniciativas interinstitucionais e ações voltadas ao incentivo à pesquisa, congregando especialistas de diversas áreas do conhecimento, vinculados a Instituições de Educação Superior, públicas e privadas, em âmbito nacional e internacional.

Seu principal objetivo é fortalecer a integração entre instituições, tanto no Brasil quanto no exterior, por meio de redes de pesquisa comprometidas com a formação continuada de profissionais da educação. Para isso, busca-se a produção e a ampla disseminação do conhecimento em distintas áreas do saber.

Expressamos nossa profunda gratidão aos autores pelo empenho, comprometimento e dedicação na concepção e finalização desta obra. Esperamos que ela se consolide como um recurso didático-pedagógico valioso, atendendo às necessidades de estudantes, docentes de todos os níveis de ensino e demais interessados na temática.

Ernane Rosa Martins

SUMÁRIO

Capítulo 01

TÉCNICAS DE APRENDIZADO POR REFORÇO PARA MELHORAR O COMPORTAMENTO ADAPTATIVO DE NPCs EM UM JOGO ROGUELIKE/ ROGUELITE

Luis Cuevas Rodríguez; Clairon Lima Pinheiro; Kelly Christiane Silva de Souza; Maria Olívia de Albuquerque Ribeiro Simão; Cristina Souza de Araújo; Jucimar Maia da Silva Junior

doi 10.37885/250118746 8

Capítulo 02

ÉTICA E CONSTRUÇÃO DE CONFIANÇA EM MODELOS DE INTELIGÊNCIA ARTIFICIAL

Lucas Cerqueira Machado Dias; Gabriela de Araújo Macedo; Laura Galvão Marques Cavalcante

doi 10.37885/241218612 30

Capítulo 03

ÉTICA E INTELIGÊNCIA ARTIFICIAL: OS DESAFIOS DO PRINCÍPIO DA EXPLICABILIDADE À LUZ DO PL 2.338/23

Gabriela de Araújo Macedo; Lucas Cerqueira Machado Dias; Laura Galvão Marques Cavalcante

doi 10.37885/250218817 45

Capítulo 04

IA Y ROBOTICS APLICADOS AL ANÁLISIS FINANCIERO EN UNA EMPRESA DEL SECTOR PESQUERO: UNA APROXIMACIÓN DESDE LA CONTABILIDAD GERENCIAL

Patricia Cáceres Ortega; Sonia Jackeline Miranda Avalos; Juan Alberto Villagómez Chinchay; Edwin Sixto Moncada Ochoa; Víctor Dante Ataupillco Vera

doi 10.37885/250318963 62

Capítulo 05

APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA OCEANOGRAFIA: INOVAÇÕES, DESAFIOS E PERSPECTIVAS

Erick Paiva Barbieri; Edison Barbieri

doi 10.37885/250419199 85

Capítulo 06**INTELIGÊNCIA ARTIFICIAL NA AQUICULTURA: CAMINHOS PARA UMA PRODUÇÃO SUSTENTÁVEL E AUTOMATIZADA**

Erick Paiva Barbieri; Edison Barbieri

doi 10.37885/250419220 105**Capítulo 07****RECONHECIMENTO FACIAL E RACISMO ALGORÍTMICO: OS DESAFIOS NA SEGURANÇA PÚBLICA BRASILEIRA**

Roberto Cezar Marcelino Neris

doi 10.37885/250519437 125**Capítulo 08****RECONHECIMENTO FACIAL E REPRESENTATIVIDADE: A INFLUÊNCIA DAS VARIÁVEIS NA PRECISÃO DOS ALGORITMOS**

Roberto Cezar Marcelino Neris

doi 10.37885/250519438 145**Capítulo 09****TRANSFORMACIÓN DIGITAL, INTELIGENCIA ARTIFICIAL Y AUTOMATIZACIÓN: RECONFIGURACIÓN DE LOS MODELOS DE GESTIÓN ORGANIZACIONAL**

Pilar Nely Flores Salazar; Wilderth Erick Velasquez Medina; Marisol Gonzaga Zirena Cano; Bertelly Turpo Aliaga; Delia Concepcion Cahuana Pacco

doi 10.37885/250619581 163**SOBRE O ORGANIZADOR** 180**ÍNDICE REMISSIVO** 181

TÉCNICAS DE APRENDIZADO POR REFORÇO PARA MELHORAR O COMPORTAMENTO ADAPTATIVO DE NPCS EM UM JOGO ROGUELIKE/ ROGUELITE

Luis Cuevas Rodríguez

Universidade do Estado do Amazonas

Clairon Lima Pinheiro

Universidade do Estado do Amazonas

Kelly Christiane Silva de Souza

Universidade do Estado do Amazonas

Maria Olívia de Albuquerque Ribeiro Simão

Universidade Federal do Amazonas

Cristina Souza de Araújo

Universidade do Estado do Amazonas

Jucimar Maia da Silva Junior

Laboratório LUDUS

Universidade do Estado do Amazonas

RESUMO

Os jogos Roguelike/Roguelite destacam-se por sua geração procedural, alta dificuldade e rejogabilidade, proporcionando experiências únicas a cada sessão. Neste contexto, os NPCs (Personagens Não-Jogáveis) desempenham um papel fundamental ao interagir dinamicamente com os jogadores. Este estudo revisa o uso de técnicas de Aprendizado por Reforço (Reinforcement Learning - RL) aplicadas ao comportamento adaptativo de NPCs em jogos Roguelike, permitindo que estes aprendam e ajustem suas estratégias com base no desempenho e nas ações dos jogadores. A adaptação dos NPCs promove uma experiência mais envolvente, desafiadora e personalizada. Além de revisar os métodos e aplicações existentes, o trabalho discute desafios, como o alto custo computacional e a necessidade de balanceamento entre usar o já conhecido ou buscar outras alternativas que podem ser melhores, apresentando também oportunidades para futuras pesquisas.

Palavras-chave: jogos roguelike/roguelite; aprendizado por reforço; NPCs adaptativos; ia; comportamento adaptativo.

INTRODUÇÃO

Os jogos Roguelike/Roguelite têm ganhado destaque no cenário atual dos jogos digitais devido à sua capacidade de oferecer experiências de jogo dinâmicas e desafiadoras. Caracterizados pela geração procedural de níveis, morte permanente e alta rejogabilidade, esses subgêneros exigem que os jogadores desenvolvam habilidades adaptativas e tomem decisões estratégicas sob pressão (Laurindo *et al.*, 2024; Minini, 2021). Um elemento central nessa experiência são os Personagens Não-Jogáveis (NPCs), que, quando projetados com comportamentos inteligentes e adaptativos, podem enriquecer a imersão e o dinamismo do jogo (Nagarkar, 2024; Sousa, 2022).

Nesse cenário, as técnicas de Inteligência Artificial (IA) desempenham um papel crucial no desenvolvimento de NPCs mais sofisticados e realistas. Em particular, o Aprendizado por Reforço (Reinforcement Learning - RL) surge como uma abordagem promissora, permitindo que os NPCs aprendam e evoluam suas estratégias com base em tentativa e erro e na interação com o ambiente (Padakandla, 2021; Sestini *et al.*, 2020a). Diferente de abordagens tradicionais, como Máquinas de Estado Finito (FSM) e Árvores de Comportamento (BTs), que resultam em comportamentos previsíveis e repetitivos, o RL permite que os NPCs desenvolvam respostas dinâmicas e adaptativas, ajustando-se às ações dos jogadores (Chen, 2024; Sukmana *et al.*, 2024).

A aplicação de algoritmos como Q-Learning, SARSA e Deep Q-Networks (DQN) em jogos Roguelike abre novas possibilidades no design de NPCs. Esses métodos possibilitam a criação de adversários ou aliados capazes de oferecer desafios únicos e equilibrar dinamicamente a dificuldade do jogo, aumentando a rejogabilidade e a imersão (Sestini *et al.*, 2020b; De Woillemont, 2023). No entanto, desafios significativos ainda persistem, como o alto custo computacional e a definição de recompensas adequadas para guiar o aprendizado dos agentes.

Este trabalho revisa o uso de técnicas de aprendizado por reforço aplicadas ao comportamento adaptativo de NPCs em jogos Roguelike, discutindo os principais algoritmos utilizados, suas vantagens, desafios e oportunidades. A análise busca identificar como o uso de RL pode transformar a experiência do jogador, promovendo interações mais dinâmicas, realistas e personalizadas.

METODOLOGIA

A pesquisa foi conduzida para identificar estudos relevantes sobre o uso de técnicas de aprendizado por reforço (Reinforcement Learning - RL) aplicadas a NPCs em jogos Roguelike/Roguelite. A seguir, detalham-se os critérios de busca, as bases de dados utilizadas e os critérios de inclusão e exclusão.

Em bases de dados acadêmicas como ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect e no Google Scholar foram utilizadas palavras-chave em inglês quanto em português para fazer busca da literatura relacionada com o tema. Os termos em inglês incluíram 'Roguelike/Roguelite games', 'AI in Roguelike/Roguelite games', 'NPC adaptive behavior' e 'Reinforcement Learning', enquanto os termos em português incluíram 'Jogos Roguelike/Roguelite', 'Inteligência Artificial em jogos Roguelike/Roguelite', 'Comportamento adaptativo de NPCs' e 'Aprendizado por Reforço'. Nas buscas foram usados os dois termos "Roguelike" e "Roguelite".

Por exemplo, para pesquisas em bases de dados bilíngues ou que suportem mais de um idioma se usaram critérios de busca como:

```
("Reinforcement Learning" OR "Aprendizado por Reforço") AND  
("NPC adaptive behavior" OR "Comportamento adaptativo de  
NPCs") AND ("Reinforcement Learning" OR "Aprendizado por  
Reforço") AND ("Roguelike" OR "Roguelite" OR "Jogos Roguelite")
```

Os filtros aplicados incluem a seleção de estudos publicados só entre 2010 e 2024, apenas artigos em inglês e em português e artigos completos sem custo para consultas.

Estratégia de Análise dos Resultados

Os artigos selecionados foram analisados qualitativamente, buscando identificar tendências, avanços e lacunas no uso de aprendizado por reforço em jogos digitais. Para isso, os resultados foram organizados em categorias que facilitassem a interpretação e a contextualização das contribuições de cada estudo.

A primeira categoria explorou o que são jogos eletrônicos tipo Roguelike/Roguelite suas características e diferenças, papel dos NPC nesse tipo de jogos e importância da adaptabilidade desses NPCs. A segunda categoria abordou a inteligência artificial para NPCs em jogos digitais e as características do

aprendizado por reforço. E uma terceira categoria centrou-se no uso do aprendizado por reforço para modelar o comportamento de NPCs em jogos Roguelike.

Nas próximas seções, serão introduzidos conceitos fundamentais sobre jogos Roguelike/Roguelite, a aplicação de inteligência artificial (IA) nesses jogos e os princípios do aprendizado por reforço. Será discutido como os jogos Roguelike/Roguelite se caracterizam por elementos como geração procedural e desafios dinâmicos, além de explorar o papel da IA em criar experiências imersivas e adaptativas para os jogadores. Em seguida, serão abordadas as técnicas de aprendizado por reforço, que permitem que personagens não-jogáveis (NPCs) se adaptem e evoluam conforme as ações do jogador, proporcionando uma experiência mais desafiadora e envolvente.

JOGOS ROGUELIKE E ROGUELITE

Os jogos Roguelike/ Roguelite têm sua denominação derivada do título *Rogue*, que foi lançado em 1980. *Rogue* estabeleceu padrões para jogos que se concentram na exploração de masmorras, utilização de técnicas de geração procedural e uma elevada dificuldade desafiadora. Os Roguelikes são frequentemente associados a uma experiência de jogo mais convencional, enquanto os Roguelites introduzem variações que tornam o gênero mais acessível e diversificado (Laurindo *et al.*, 2024; Sousa, 2022). Esses jogos se caracterizam por usar geração procedural, morte permanente, elevada dificuldade, progressão baseada em itens ou experiência e narrativa Emergente.

Jogos que pertencem a este gênero frequentemente exigem do jogador habilidades estratégicas aguçadas e bem desenvolvidas, além de uma rápida e eficiente capacidade de tomada de decisão sob pressão. Esses jogos desafiam os jogadores de forma contínua e intensa, proporcionando uma experiência de jogo que eleva o envolvimento a um novo nível de complexidade e profundidade (Rosa, 2022). Cada movimento, cada escolha, se torna crucial, inserindo o jogador em um estado de constante alerta e atenção, onde a estratégia adequada pode ser a diferença entre o sucesso e a falha.

A dinâmica interação entre elementos aleatórios, variáveis e imprevistos produz narrativas que se tornam singularmente envolventes e marcantes para cada jogador. Uma característica não apenas apreciada, mas também

frequentemente desejada dentro deste gênero de jogo eletrônico (Sousa, 2022). Essa singularidade confere ao jogo uma diversidade de experiências que se adaptam a cada jogador, tornando cada partida única e memorável.

Os jogos Roguelike e Roguelite compartilham elementos centrais, como mapas gerados proceduralmente, desafios progressivos e foco na rejogabilidade, mas diferem em aspectos fundamentais. Enquanto os Roguelikes seguem a rigorosa definição da "Berlin Interpretation", com combate baseado em turnos, mapas em grade e ausência de progresso entre as sessões, os Roguelites flexibilizam essas convenções, permitindo progresso persistente e introduzindo mecânicas em tempo real. Essa abordagem mais acessível torna os Roguelites atrativos para um público diversificado (Laurindo *et al.*, 2024; Minini, 2021).

Nesta pesquisa, será adotado o termo Roguelike para referenciar ambos os subgêneros, considerando que em ambos a adaptabilidade dos NPCs é um elemento de grande relevância. A introdução de NPCs adaptativos tem o potencial de enriquecer a experiência do jogador, aumentando a imersão, a estratégia e a rejogabilidade, independentemente das especificidades de cada subgênero.

Personagens Não-Jogáveis (NPCs) e seu Papel nos Jogos Digitais

Os Personagens Não-Jogáveis (NPCs) estão presentes em boa parte dos jogos digitais e desempenham papéis variados que vão desde aliados e inimigos a fornecedores de missões especiais. Além disso, eles podem atuar como guias, auxiliando o jogador a entender o que fazer em momentos específicos do jogo. Dessa forma, os NPCs contribuem de maneira significativa para enriquecer a experiência de jogo, seja ajudando o jogador a compreender melhor o mundo narrativo, seja oferecendo ferramentas ou itens necessários para superar os desafios da jogabilidade. Por essa razão, os NPCs desempenham um papel vital no design de jogos ao alinhar elementos narrativos e de jogabilidade, criando uma experiência mais coesa e envolvente para o jogador (Nagarkar, 2024; Uludağlı; Oğuz, 2023).

A criação de NPCs que atuem de maneira realista e crível é uma tarefa desafiadora. De acordo com algumas correntes de pensamento, os NPCs podem se destacar tanto em termos de personalidade e importância para a jogabilidade que acabam assumindo um papel central no jogo, enquanto outros podem se limitar a funções secundárias. Em ambos os casos, os NPCs enriquecem o

jogo, não apenas contribuindo para a narrativa, mas também adicionando uma camada de detalhamento e interatividade ao ambiente, especialmente em jogos que exigem escolhas estratégicas e emocionais (Nagarkar, 2024).

Finalmente, os NPCs não apenas ajudam a criar um ambiente de jogo mais realista, mas também servem como elementos para modular a experiência do jogador. Eles podem atuar como inimigos reais, interrompendo o progresso do jogador e criando desafios que precisam ser superados para avançar no jogo. Essa complexidade torna os NPCs componentes indispensáveis no design de jogos modernos e destaca sua relevância no estudo de interatividade e inteligência artificial em jogos digitais (Uludağlı; Oğuz, 2023).

Adaptabilidade dos NPCs em jogos Roguelike

A inclusão de NPCs com comportamento adaptativo pode redefinir a experiência de jogo em Roguelikes e Roguelites, contribuindo para a complexidade e a profundidade desses subgêneros. Em jogos baseados na geração procedural, como os Roguelikes, NPCs adaptativos adicionam uma camada de imprevisibilidade, uma vez que aprendem e evoluem com as ações do jogador, criando interações únicas e desafiadoras. Essa capacidade de adaptação é essencial para reforçar a rejogabilidade, característica intrínseca do gênero (Silva, 2024; Vodák, s.d.).

No contexto dos Roguelikes, a dificuldade elevada é frequentemente uma marca registrada, o que pode gerar tanto frustrações quanto recompensas significativas. NPCs que ajustam suas estratégias com base no desempenho do jogador ajudam a equilibrar a experiência, reduzindo barreiras para jogadores novatos sem comprometer o desafio para jogadores mais experientes. Esse tipo de ajuste dinâmico promove uma curva de aprendizado mais fluida, permitindo que o jogador se sinta constantemente desafiado de forma justa (Rosa, 2022; Amenyro, 2022).

Além disso, NPCs adaptativos podem transformar a narrativa emergente desses jogos. Como Palmer (2020) destaca, a interação entre o jogador e o ambiente, incluindo NPCs, é uma das principais fontes de narrativa em jogos de exploração de masmorras. Quando os NPCs são capazes de reagir de maneira única às escolhas do jogador, eles não apenas enriquecem a história, mas também reforçam a imersão, oferecendo ao jogador a sensação de estar em um

mundo vivo e dinâmico. Essa interação dinâmica é particularmente importante em jogos onde a exploração procedural pode, por vezes, parecer repetitiva.

Nos Roguelites, a adaptabilidade dos NPCs pode desempenhar um papel ainda mais significativo, dada a progressão persistente que caracteriza o subgênero. Esses NPCs podem evoluir ao longo das sessões, refletindo as escolhas e habilidades acumuladas pelo jogador. Isso cria uma variedade contínua de desafios e incentiva o engajamento a longo prazo, especialmente em jogos com ciclos narrativos mais curtos e recompensas mais imediatas (Premoli; Silva, 2024; Wolff, 2024).

Por fim, ao integrar sistemas adaptativos em NPCs, os desenvolvedores têm a oportunidade de explorar novas possibilidades de design que maximizem o impacto emocional e estratégico para os jogadores. Essa abordagem pode redefinir as convenções do gênero, tornando os NPCs não apenas adversários ou aliados, mas parte integrante da construção da experiência narrativa e da mecânica de jogo. Assim, a adaptabilidade dos NPCs representa uma ferramenta poderosa para enriquecer Roguelikes e Roguelites, conectando os jogadores de forma mais profunda ao mundo virtual que exploram.

INTELIGÊNCIA ARTIFICIAL PARA NPC EM JOGOS DIGITAIS

A inteligência artificial (IA) é um subcampo fascinante da ciência da computação que busca modelar e, em certos casos, reproduzir o comportamento inteligente em diversos sistemas. Essa área tem como principal foco o desenvolvimento de agentes inteligentes, ou seja, entidades que atuam de maneira autônoma e eficiente em seus ambientes (Gao *et al.*, 2024).

Quando aplicada ao universo dos jogos digitais, a IA assume um papel particular, concentrando-se na reprodução precisa do comportamento de personagens não-jogáveis (NPCs) dentro do mundo virtual. Alimentados por tecnologias avançadas, os NPCs são projetados para simular comportamentos semelhantes aos humanos, oferecendo aos jogadores uma experiência mais imersiva e interativa (Karaca *et al.*, 2023).

Entre suas múltiplas funções, a IA é capaz de gerenciar interações em tempo real, aprender com as ações dos jogadores, resolver problemas dentro do jogo, tomar decisões de maneira racional e até realizar processamento de

linguagem natural para interagir verbalmente com os jogadores. No entanto, apesar de suas capacidades notáveis, a IA nos jogos ainda apresenta limitações, especialmente no que diz respeito à simulação de atividades cerebrais complexas, um obstáculo que reflete as lacunas no conhecimento atual das neurociências e ciências cognitivas. Esse cenário destaca a contínua necessidade de pesquisa e desenvolvimento para alcançar comportamentos mais realistas e avançados tanto nos jogos quanto em aplicações externas (Gao *et al.*, 2024; Karaca *et al.*, 2023).

No contexto dos jogos digitais, a IA é amplamente utilizada para criar NPCs ou agentes que apresentam comportamentos com níveis variados de autonomia. Esses agentes podem responder a estímulos específicos ou até mesmo se comunicar entre si por meio de sistemas baseados em regras, lógica difusa (fuzzy logic) ou aprendizado por reforço. De forma geral, a IA é categorizada em três grandes tipos: reativa, subordinada e autônoma. Embora o termo IA seja frequentemente utilizado para descrever agentes com uma ou mais dessas características, nesta discussão ele será restrito aos agentes reativos e aqueles com capacidade de aprendizado. Agentes reativos tomam decisões com base em estímulos atuais, utilizando processos lógicos como árvores de decisão e sistemas baseados em regras. No entanto, esses agentes não fazem considerações prévias ou futuras, limitando-se ao presente contexto do jogo (Hauptman *et al.*, 2023).

Por outro lado, agentes subordinados tomam decisões baseadas em experiências, conhecimentos e regras fixas, representadas por redes neurais, sistemas especialistas ou abordagens baseadas em condições comportamentais. Esses agentes possuem uma certa capacidade de comunicação entre si, permitindo que atualizem protocolos e objetivos comuns conforme progridem. Ainda assim, esse tipo de IA é limitado pela experiência singular do desenvolvedor no momento da criação, tornando-se rígido e incapaz de se adaptar a novas versões ou mudanças no algoritmo do jogo (Karaca *et al.*, 2023).

Para superar essas limitações, a IA dinâmica é amplamente explorada no desenvolvimento de jogos, com sistemas que utilizam lógica difusa, redes neurais e árvores de decisão cujos pesos são ajustados dinamicamente em cenários não previstos. Esses métodos permitem que os NPCs apresentem comportamentos adaptáveis e evoluam conforme a situação do jogo, oferecendo uma experiência

mais realista e desafiadora ao jogador. Em teoria, essa abordagem supera as limitações das IAs reativas e subordinadas, permitindo um comportamento mais próximo do ideal para agentes autônomos em ambientes virtuais complexos (Hauptman *et al.*, 2023).

Por fim, a integração de sistemas de inteligência artificial no comportamento dos NPCs permite a criação de personagens que não apenas reagem ao ambiente, mas também evoluem conforme interagem com os jogadores e com o mundo virtual. Essa evolução é especialmente relevante quando abordamos técnicas que possibilitam comportamentos adaptativos e inteligentes, como o aprendizado por reforço. Nesse contexto, os agentes conseguem aprender a partir de tentativas e erros, ajustando suas ações para alcançar objetivos específicos dentro do jogo. A aplicação dessas técnicas representa um avanço significativo no desenvolvimento de NPCs, criando experiências dinâmicas, desafiadoras e mais imersivas.

Na próxima seção, se discutem os conceitos fundamentais do aprendizado por reforço, explorando suas bases teóricas, funcionamento e aplicabilidade no contexto dos jogos digitais, especialmente no desenvolvimento de NPCs capazes de se adaptar e aprender com as interações dos jogadores.

APRENDIZADO POR REFORÇO

O Aprendizado por Reforço (Reinforcement Learning - RL) é um paradigma essencial na inteligência artificial voltado para a criação de sistemas autônomos capazes de aprender a tomar decisões ótimas por meio da interação com o ambiente. Diferente do aprendizado supervisionado ou não-supervisionado, o RL se baseia em um ciclo de tentativa e erro em que um agente executa ações para maximizar a soma de recompensas cumulativas ao longo do tempo. Essa abordagem enfatiza a necessidade de equilibrar exploração (descobrir novas estratégias e possibilidades) e exploração (utilizar o conhecimento já adquirido para obter recompensas imediatas) como um desafio central do processo de aprendizado (Padakandla, 2021; Gu *et al.*, 2024).

Uma das primeiras contribuições relevantes para esta discussão é o trabalho de (Alves da Silva Lopes; Guerresi de Mello Braga, 2017), que introduzem a aprendizagem por reforço como um método em que agentes de software tomam

decisões em um ambiente com o objetivo de maximizar recompensas cumulativas. Os autores ressaltam que, ao contrário do aprendizado supervisionado, não são fornecidos pares de entrada/saída corretos, o que torna o aprendizado por reforço uma abordagem única e desafiadora.

Em seguida, (Roberto Caixeta Filho, 2018) aprofunda a discussão, apresentando o aprendizado por reforço como uma abordagem de comportamento de máquina que utiliza a interação com um ambiente dinâmico por meio de tentativa e erro. O autor destaca a importância das atualizações de ações baseadas em recompensas, enfatizando que o agente deve descobrir uma política eficaz a partir de suas experiências, sem conhecimento prévio sobre o ambiente. A introdução do conceito de trade-off entre exploração e exploração é um ponto crítico, pois reflete os desafios que os agentes enfrentam ao tentar equilibrar a busca por novas informações com a utilização do conhecimento já adquirido.

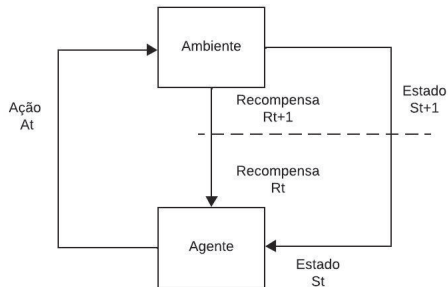
Por fim, (da Costa Mota, 2018) amplia a compreensão do aprendizado por reforço ao contextualizá-lo no âmbito da inteligência artificial e da robótica. O autor classifica a aprendizagem de máquina em três categorias, incluindo a aprendizagem por reforço, e destaca a interação do agente com um ambiente dinâmico como essencial para a resolução de objetivos. Essa perspectiva reforça a ideia de que o aprendizado por reforço não é apenas uma técnica, mas um campo de estudo fundamental para o desenvolvimento de sistemas autônomos que aprendem e se adaptam a partir da experiência.

No aprendizado por reforço, o agente (o tomador de decisão) interage com um ambiente dinâmico em que observa um estado, realiza uma ação e recebe uma recompensa numérica em resposta. Esse ciclo se repete até que o agente atinja um estado terminal. O objetivo é que o agente aprenda qual ação maximiza a recompensa acumulada ao longo de suas interações. Esse processo é frequentemente modelado como um Processo de Decisão de Markov (MDP), Figura 1, composto pelos seguintes elementos:

- Estados (S): Representações do ambiente em um determinado instante.
- Ações (A): Decisões que o agente pode tomar.
- Recompensas (R): Valores numéricos que indicam o sucesso de uma ação.

- Modelo de Transição: Probabilidade de mudar de um estado para outro ao tomar uma ação.
- Política (π): Estratégia que define as ações que o agente escolhe para maximizar suas recompensas (Canese *et al.*, 2021; Nguyen *et al.*, 2020).

Figura 1 - Representação clássica do Processo de Decisão de Markov (MDP) (autor).



Na Figura 1 descreve a interação entre um agente e um ambiente em uma sequência de decisões ao longo do tempo. O agente observa o estado atual (S_t) do ambiente e, com base nessa observação, seleciona uma ação (A_t). Essa ação é enviada ao ambiente, que responde fornecendo dois elementos: uma recompensa (R_t), um valor numérico que sinaliza a quão boa foi a ação tomada, e um novo estado (S_{t+1}), que representa a situação atualizada do ambiente após a ação do agente. Esse ciclo de decisão continua de forma iterativa, com o objetivo do agente sendo aprender uma política ótima, ou seja, uma sequência de ações que maximiza a recompensa acumulada ao longo do tempo. O MDP é essencial no aprendizado por reforço, pois assume que o ambiente segue a propriedade de Markov, onde o próximo estado (S_{t+1}) e a recompensa (R_{t+1}) dependem exclusivamente do estado atual (S_t) e da ação escolhida (A_t), independentemente do histórico passado.

O aprendizado ocorre ao longo de um trade-off entre exploração e exploração, em que o agente busca equilibrar a tentativa de descobrir novas ações e a execução das melhores ações já conhecidas. O conceito de função de valor é fundamental no RL, pois estima a recompensa total esperada a partir de um estado, enquanto a função Q estende essa ideia para ações específicas em um determinado estado.

Tipos de Algoritmos de Aprendizado por Reforço

Os algoritmos de RL podem ser amplamente classificados em dois tipos principais: baseados em valor e baseados em política.

Métodos Baseados em Valor

Os métodos baseados em valor se concentram no aprendizado da função de valor, que reflete o retorno esperado ao seguir uma determinada política. O algoritmo Q-Learning é um exemplo clássico desse tipo, pois permite que o agente aprenda uma política ótima mesmo sem um modelo do ambiente, utilizando um processo de tentativa e erro. Além disso, métodos como o aprendizado por diferença temporal (TD-Learning) permitem atualizar a função de valor com base em experiências parciais, reduzindo o custo computacional (Padakandla, 2021).

Para resolver o problema do espaço de estado-ação ser muito grande, métodos de aproximação de função são utilizados, como redes neurais profundas no Deep Q-Learning, permitindo a generalização em larga escala. A Tabela 1 resume os principais métodos baseados em valor.

Tabela 1 - Métodos Baseados em Valor.

Método	Descrição	Aplicações Comuns
Q-Learning	Aprende a função Q sem modelo do ambiente	Jogos, robótica
SARSA	Método on-policy com atualização TD	Controle em tempo real
Deep Q-Learning	Generaliza com redes neurais	Jogos complexos

Métodos Baseados em Política

Os métodos baseados em política focam diretamente no aprendizado da política ótima. Em vez de estimar valores para estados ou ações, esses algoritmos ajustam a política com base em gradientes de política. Eles são particularmente eficazes em ambientes com ações contínuas, onde a discretização seria impraticável (Gu *et al.*, 2024).

O aprendizado por reforço tem sido amplamente aplicado em diversas áreas, incluindo robótica, finanças, saúde e, especialmente, jogos digitais. Nos jogos, o RL tem sido responsável por avanços notáveis no desenvolvimento de

NPCs inteligentes, capazes de aprender estratégias complexas e superar até mesmo adversários humanos. Isso foi exemplificado no treinamento de agentes como AlphaGo e AlphaStar, que atingiram níveis super-humanos em jogos como Go e StarCraft II.

O uso de RL em jogos se destaca devido ao ambiente bem-definido e à capacidade de gerar grandes volumes de dados simulados a baixo custo. Além disso, o RL permite a criação de NPCs adaptativos que ajustam seu comportamento com base nas ações do jogador, tornando a experiência de jogo dinâmica e desafiadora (Nguyen *et al.*, 2020; Canese *et al.*, 2021).

O aprendizado por reforço representa um avanço fundamental na criação de sistemas inteligentes capazes de tomar decisões ótimas por meio da interação com o ambiente. Sua aplicação em jogos digitais, especialmente no desenvolvimento de NPCs, destaca o potencial dessa abordagem para oferecer experiências mais imprevisíveis, imersivas e realistas aos jogadores. Na próxima seção, exploraremos algoritmos específicos e suas aplicações no comportamento adaptativo de NPCs em jogos Roguelike/Roguelite.

APRENDIZADO POR REFORÇO APLICADO AO COMPORTAMENTO DE NPCs EM JOGOS ROGUELIKE

O uso de Aprendizado por Reforço (Reinforcement Learning - RL) no comportamento de Personagens Não Jogadores (NPCs) tem emergido como uma abordagem promissora para aprimorar a experiência dos jogadores em jogos Roguelike. Esses jogos, conhecidos por sua geração procedural e altos níveis de desafio, exigem NPCs dinâmicos e adaptativos que sejam capazes de responder às ações dos jogadores de forma inteligente. A aplicação de técnicas de RL possibilita que os NPCs aprendam comportamentos complexos e evoluam suas estratégias, tornando as interações mais imprevisíveis e imersivas (Sestini *et al.*, 2020a; De Woillemont, 2023).

Um dos principais benefícios do aprendizado por reforço no desenvolvimento de NPCs é a capacidade de superar limitações tradicionais encontradas em algoritmos baseados em árvores de comportamento (BTs) ou máquinas de estado finito (FSM). Conforme destacado por Sestini *et al.* (2020a), abordagens convencionais muitas vezes resultam em comportamentos previsíveis e repetitivos, limitando o realismo e a profundidade das interações. Com o RL,

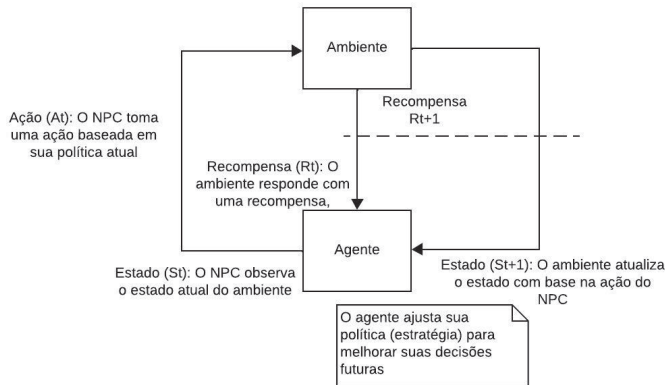
os NPCs conseguem tomar decisões em tempo real com base em feedbacks adaptativos do ambiente, explorando novas estratégias enquanto otimizam suas recompensas acumuladas. Isso é particularmente relevante em jogos Roguelike, onde a natureza procedimental dos cenários exige que os NPCs se ajustem a contextos inesperados e dinâmicos (Chen, 2024).

Para abordar as deficiências observadas em métodos tradicionais, modelos avançados têm sido implementados utilizando políticas profundas e técnicas de RL baseadas em valor e baseadas em política. Por exemplo, métodos como Q-Learning e Deep Q-Networks (DQN) permitem que os NPCs aprendam políticas ótimas sem a necessidade de um modelo explícito do ambiente. Em contrapartida, algoritmos baseados em gradiente de política possibilitam tomadas de decisão em espaços contínuos, sendo particularmente úteis para simulações complexas em Roguelikes (Sukmana *et al.*, 2024; Canese *et al.*, 2021).

Além de resolver limitações comportamentais, o RL também pode ser utilizado para ajustar dinamicamente a dificuldade do jogo. Segundo Chen (2024), essa técnica, conhecida como Dynamic Difficulty Adjustment (DDA), permite que os NPCs modifiquem seu comportamento com base no desempenho do jogador. Dessa forma, o jogo mantém um equilíbrio entre desafio e recompensa, oferecendo uma experiência personalizada e engajante. Esse tipo de adaptação é essencial em jogos Roguelike, onde a dificuldade elevada é uma característica central, mas que, quando mal dosada, pode frustrar jogadores menos experientes.

A Figura 2 a seguir ilustra o ciclo de aprendizado de um NPC em um ambiente de jogo Roguelike, baseado no processo de decisão de Markov (MDP). Nesse ciclo, o agente (NPC) observa o estado do ambiente, realiza uma ação e recebe uma recompensa, ajustando sua política ao longo do tempo.

Figura 2 - NPC em um jogo Roguelike, baseado no processo de decisão de Markov (MDP) (autor).



O diagrama da Figura 2 apresenta duas entidades principais: o agente (NPC) e o ambiente, que interagem de forma contínua. O ciclo inicia quando o agente observa o estado atual (S_t) do ambiente, representando a situação em que ele se encontra no momento. Com base nesse estado e em sua política atual, o agente toma uma ação (A_t), que influencia diretamente o ambiente. Em resposta, o ambiente fornece uma recompensa (R_t), indicando o sucesso ou fracasso da ação tomada, e atualiza o estado para um novo valor (S_{t+1}), refletindo as mudanças causadas pela ação. O agente, por sua vez, utiliza as informações da recompensa e do novo estado para ajustar sua política, buscando melhorar suas decisões futuras. Esse processo iterativo se repete continuamente, permitindo que o agente aprenda de forma autônoma, adaptando seu comportamento ao ambiente dinâmico do jogo. Dessa forma, o ciclo ilustra como os NPCs em jogos Roguelike podem evoluir suas estratégias com base em tentativas e erros, maximizando recompensas ao longo do tempo e proporcionando uma experiência de jogo mais desafiadora e imprevisível.

Na próxima seção, se apresentarão três dos principais algoritmos de aprendizado por reforço utilizados para o comportamento de NPCs em jogos Roguelike: Q-Learning, SARSA e Deep Q-Networks (DQN), destacando suas características, diferenças e aplicações no contexto de ambientes dinâmicos e procedurais

Algoritmos de Aprendizado por Reforço Aplicados ao Comportamento de NPCs Q-Learning

Os algoritmos Q-Learning, SARSA e Deep Q-Networks (DQN) são algoritmos que utilizam o aprendizado por reforço (Reinforcement Learning - RL). Eles são classificados como métodos baseados em valor, ou seja, focam no aprendizado de funções de valor que permitem ao agente determinar a melhor ação a ser tomada em um determinado estado para maximizar a recompensa cumulativa.

O Q-Learning é um algoritmo baseado em valor, amplamente utilizado no aprendizado por reforço. Ele permite que o agente aprenda uma política ótima sem precisar de um modelo do ambiente. O Q-Learning atualiza os valores de Q (ação-estado) iterativamente com base na equação de Bellman, utilizando uma abordagem off-policy, ou seja, o agente pode aprender uma política ótima mesmo enquanto explora outras ações (Sestini *et al.*, 2020b). No contexto de NPCs em jogos Roguelike, o Q-Learning permite que os agentes aprendam comportamentos adaptativos ao longo do tempo, ajustando suas estratégias com base nas recompensas obtidas em diferentes estados do jogo.

O algoritmo SARSA é uma variação do Q-Learning, mas utiliza uma abordagem on-policy, onde a política seguida pelo agente é a mesma que está sendo aprendida. O nome SARSA é uma abreviação dos elementos envolvidos: State-Action-Reward-State-Action. Esse método é mais estável em ambientes onde a exploração é perigosa, sendo adequado para NPCs que precisam manter comportamentos mais controlados, como evitar decisões arriscadas que possam prejudicar o jogador ou o fluxo do jogo (Lie; Istiono, 2022).

O DQN é uma extensão do Q-Learning que utiliza redes neurais profundas para aproximar a função Q em ambientes com espaços de estado contínuos ou muito grandes. Desenvolvido inicialmente para jogos complexos, como o Atari, o DQN se mostrou altamente eficaz para treinar NPCs que precisam lidar com cenários dinâmicos e não estruturados, como nos jogos Roguelike (Sestini *et al.*, 2020b; Sukmana *et al.*, 2024). Esse método resolve o problema de escalabilidade do Q-Learning tradicional e permite que os NPCs desenvolvam comportamentos sofisticados e adaptativos, aumentando a imprevisibilidade e a rejogabilidade do jogo.

A Tabela 2 fornece uma visão comparativa objetiva dos métodos de aprendizado por reforço discutidos na seção, facilitando a compreensão dos leitores sobre a escolha e aplicação desses algoritmos no comportamento de NPCs em jogos Roguelike.

Tabela 2 - Comparação entre métodos de RL: Q-Learning, SARSA e DQN.

Algoritmo	Desempenho	Custo Computacional	Complexidade Comportamental	Referências
Q-Learning	Alto desempenho em problemas com poucos estados.	Baixo em ambientes simples; cresce com estados.	Baixa: comportamentos discretos e limitados.	Sestini <i>et al.</i> (2020a), Lie; Istiono (2022)
SARSA	Desempenho estável em cenários com risco controlado.	Similar ao Q-Learning, mas mais conservador.	Moderada: evita comportamentos arriscados.	Sukmana <i>et al.</i> (2024), Canese <i>et al.</i> (2021)
DQN	Alto desempenho em espaços de estado contínuos.	Alto: redes neurais exigem recursos robustos.	Alta: comportamento mais sofisticado e adaptativo.	Sestini <i>et al.</i> (2020a), Chen (2024)

Apesar de suas vantagens, a aplicação de RL em jogos Roguelike não está isenta de desafios. Conforme discutido por Tilak (2024), a eficiência amostral ainda é uma limitação significativa, exigindo grandes quantidades de dados e poder computacional para treinar agentes de maneira eficaz. Além disso, a definição de funções de recompensa adequadas é crucial, pois recompensas mal projetadas podem resultar em comportamentos indesejados ou não alinhados aos objetivos do jogo. Esses desafios destacam a necessidade de um equilíbrio cuidadoso entre exploração e exploração durante o processo de treinamento dos NPCs (Lai, 2024; Lie; Istiono, 2022).

Em resumo, o aprendizado por reforço oferece uma estrutura robusta para o desenvolvimento de NPCs adaptativos em jogos Roguelike, permitindo interações mais dinâmicas e imersivas. Ao superar as limitações dos métodos tradicionais e introduzir comportamentos baseados em aprendizado contínuo, o RL não apenas enriquece a experiência do jogador, mas também impulsiona a inovação no design de jogos procedurais. Contudo, questões relacionadas à eficiência computacional e ao design das funções de recompensa devem continuar sendo exploradas para maximizar o potencial dessa abordagem. A Tabela 1 abaixo compara os métodos tradicionais e os métodos baseados em RL aplicados ao comportamento de NPCs em Roguelikes.

Tabela 3 - Comparação entre métodos tradicionais e os métodos baseados em RL.

Método	Vantagens	Limitações
Máquinas de Estado Finito	Simple e de fácil implementação	Comportamento previsível e limitado
Árvores de Comportamento	Modularidade e controle mais refinados	Falta de adaptabilidade em cenários dinâmicos
RL (Q-Learning, DQN)	Adaptabilidade, aprendizado autônomo	Alta demanda de recursos computacionais
Gradiente de Política	Decisões em espaços contínuos	Complexidade no ajuste de recompensas

A implementação contínua de novas técnicas de RL e suas aplicações práticas em Roguelikes promete resultados ainda mais expressivos, com NPCs capazes de oferecer desafios estratégicos e narrativas emergentes cada vez mais envolventes.

CONCLUSÕES, DESAFIOS E OPORTUNIDADES

O uso de técnicas de Aprendizado por Reforço (RL) no comportamento de NPCs adaptativos em jogos Roguelike apresentou avanços significativos no desenvolvimento de personagens não-jogáveis mais inteligentes e responsivos. Os métodos empregados, como Q-Learning, SARSA e Deep Q-Networks (DQN), demonstraram a capacidade de criar comportamentos únicos e adaptáveis, proporcionando uma experiência mais dinâmica e imersiva para os jogadores. A análise das métricas de desempenho revelou que a evolução comportamental dos NPCs pode ser medida não apenas pela precisão, mas pelo número de padrões de comportamento únicos que emergem no ambiente de jogo. Entretanto, algumas limitações foram identificadas, como taxas de convergência lentas, uso inadequado de conhecimento inicial e pequenos problemas na exploração durante as fases iniciais do aprendizado (Sestini *et al.*, 2020b; De Woillemont, 2023).

Apesar dos avanços, constatou-se que o RL por si só não é suficiente para lidar com todos os desafios apresentados no desenvolvimento de NPCs. Em algumas situações, foi necessário recorrer a técnicas complementares, como Aprendizado por Reforço Profundo (DRL) ou métodos de planejamento tradicionais, como Goal-Oriented Action Planning (GOAP) e Monte Carlo Tree Search (MCTS). A combinação de múltiplas abordagens (híbridas) mostrou ser promissora, permitindo a resolução simultânea de diferentes problemas de

design e melhorando a capacidade dos NPCs de se adaptar a cenários complexos. Além disso, observou-se que o sucesso dessas técnicas depende não apenas da qualidade dos algoritmos, mas também de ajustes no próprio design do jogo. A colaboração entre pesquisadores de IA e designers de jogos se mostrou fundamental para integrar com sucesso NPCs adaptativos em ambientes procedurais como os jogos Roguelike (Sukmana *et al.*, 2024; Tilak, 2024).

O estudo também identificou oportunidades importantes relacionadas ao papel dos NPCs no enriquecimento da experiência do jogador. Ao implementar NPCs com aprendizado adaptativo, é possível criar sistemas que simulam interações humanas de maneira mais convincente, oferecendo adversários inteligentes ou até aliados que evoluem conforme a progressão do jogo. Isso amplia as possibilidades de narrativa emergente, onde as decisões dos jogadores influenciam diretamente o comportamento e a estratégia dos NPCs, criando cenários únicos e imprevisíveis em cada sessão de jogo. Essa dinâmica melhora a longevidade e a replayability, aspectos fundamentais em jogos Roguelike e Roguelite (Chen, 2024; Lai, 2024).

No entanto, desafios significativos ainda persistem. A principal limitação identificada é o alto custo computacional e a baixa eficiência amostral das técnicas de RL, que exigem uma grande quantidade de dados e recursos para treinar agentes de forma eficaz. Em jogos com orçamentos mais extensos e complexos, o desafio de escolher os algoritmos ideais e ajustá-los para diferentes comportamentos de NPCs torna-se ainda mais crítico. Além disso, o desenvolvimento de comportamentos mais complexos para NPCs, como estratégias de combate avançadas, soluções de puzzles ou interações sociais mais profundas, representa uma área promissora para pesquisas futuras (Lie; Istiono, 2022; Amenyro, 2022).

Finalmente, as oportunidades apresentadas pelo Aprendizado por Reforço no design de NPCs em jogos Roguelike são vastas. Novas pesquisas devem explorar a combinação de RL com técnicas híbridas, além de investigar maneiras de otimizar o uso de recursos computacionais. Além disso, ampliar o escopo dos NPCs para além das missões básicas, permitindo que realizem atividades mais orgânicas, como comércio, crafting ou suporte em combate, pode enriquecer ainda mais a imersão e a autenticidade dos jogos. À medida que a inteligência artificial avança, a colaboração entre desenvolvedores e pesquisadores será

essencial para integrar sistemas de aprendizado que equilibrem o desempenho técnico e a experiência do jogador, pavimentando o caminho para NPCs verdadeiramente adaptativos e inteligentes nos jogos do futuro.

REFERÊNCIAS

ALVES da Silva Lopes, R. & Guerese de Mello Braga, V. (2017). Um sistema para o aprendizado automático de jogos eletrônicos baseado em redes neurais e Q-Learning usando interface natural.

AMENEYRO, F. V. (2022). Analysing the Impacts of Dynamically Evolving Selection Policies in Monte Carlo Tree Search through Evolutionary Algorithms.

AYDIN, A. M., Kılıç, H., & Güran, A. (2023). A player reputation system based on belief formation among Non-Player character societies in Open-World Role-Playing games. **Entertainment Computing**.

CANESE, L., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., & Spanò, S. (2021). Multi-agent reinforcement learning: A review of challenges and applications. **Applied Sciences**, 11(11), 4948.

CHEN, S. (2024). Literature review of application of AI in improving gaming experience: NPC behavior and dynamic difficulty adjustment.

DA COSTA MOTA, Ícaro (2018). Aprendizagem por reforço utilizando Q-Learning e redes neurais artificiais em jogos eletrônicos.

DE WOILLEMONT, P. L. P. (2023). Simulation Of Diverse Human Play-Styles *In: Video Games: A Reinforcement Learning Approach*.

GAO, Y., Hu, Z., & Liu, J. (2024, September). The Potential and Mechanism of Artificial Intelligence Driven Non-Player Characters in Video Games for Anxiety Management. *In: 2024, International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, (pp. 353-362). Atlantis Press.

GU, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J., & Knoll, A. (2024). A Review of Safe Reinforcement Learning: Methods, Theories and Applications. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. [PDF]

HAUPTMAN, A. I., Schelble, B. G., McNeese, N. J., & Madathil, K. C. (2023). Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming. **Computers in Human Behavior**, 138, 107451.

KARACA, Y., Derias, D., & Sarsar, G. (2023). **AI-Powered Procedural Content Generation: Enhancing NPC Behaviour for an Immersive Gaming Experience**. Available at SSRN 4663382.

LAI, B. (2024). **Learning by Playing: A Stock Market Simulation Game With Deep Reinforcement Learning-powered NPCs**.

LAURINDO, E. K., OTI, K. H., BROCATTO, M. A., & BELCHIOR, V. (2024). Relatório técnico de desenvolvimento do jogo digital Thorn Ascent.

- LIE, C. S. K. & Istiono, W. (2022). How to make npc learn the strategy in fighting games using adaptive AI?. *Int J Sci Tech Res Eng*.
- MININI, P. P. (2021). Exploração de técnicas atuais de programação de jogos e geração procedural de ambientes através do desenvolvimento de um protótipo de jogo em Rust.
- NAGARKAR, S. B. (2024). **Improving realism and interactivity in games: a study of AI-integrated non player characters (NPCs)**.
- NGUYEN, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE transactions on cybernetics*, 50(9), 3826-3839.
- PADAKANDLA, S. (2021). A survey of reinforcement learning algorithms for dynamically varying environments. *ACM Computing Surveys (CSUR)*.
- PALMER, Z. (2020). **The Living Dungeon: Space, Convention, and Reinvention in Dungeon Games**.
- PREMOLI, B. O. & Silva, V. A. (2024). Estudo sobre o comportamento de NPCS e inimigos em jogos digitais utilizando a inteligência artificial.
- ROBERTO Caixeta Filho, E. (2018). Aprendizado por reforço aplicado ao ambiente Toy Text da plataforma Gym.
- ROSA, M. P. C. (2022). Ajuste Dinâmico de Dificuldade pelo desempenho e perfil de jogador em jogo de plataforma.
- SESTINI, A., Kuhnle, A., & Bagdanov, A. D. (2020a). Deep policy networks for NPC behaviors that adapt to changing design parameters in roguelike games. *arXiv preprint arXiv:2012.03532*.
- SESTINI, A., Kuhnle, A., & Bagdanov, A. D. (2020b). Deepcrawl: Deep reinforcement learning for turn-based strategy games. *arXiv preprint arXiv:2012.01914*.
- SILVA, J. A. (2024). Aprendizado por reforço no ambiente de jogos. repositorio.ucs.br. <https://repositorio.ucs.br/xmlui/bitstream/handle/11338/13649/TCC%20Jonatan%20Amaral%20da%20Silva.pdf>.
- SOUSA, K. A. (2022). Um estudo sobre o uso e classificação de aleatoriedade em jogos de estratégia. https://repositorio.ufc.br/bitstream/riufc/68426/3/2022_tcc_kasousa.pdf.
- SUKMANA, S. E., Khairy, M. S., Ratsanjani, M. H., & Rahmad, C. (2024, September). Reinforcement Learning for AI NPC Literacy Educational Game. *In: 2024, International Conference on Electrical and Information Technology (IEIT)*, (pp. 158-162). IEEE.
- TILAK, G. (2024). AI-DRIVEN NPCS AND THE EVOLUTION OF INTERACTIVE STORY TELLING IN VIDEO GAMES.
- ULUDAĞLI, M. & Oğuz, K. (2023). **Non-player character decision-making in computer games**. *Artificial Intelligence Review*.
- VODÁK, J. (s.d.). 2D ROGUE-LIKE GAME WITH PROCEDURAL ELEMENTS. *theses.cz*.
- WOLFF, A. (2024). **Building a Legacy: An Autoethnographic Reflection on Devising Theatre of Hopology**.

ÉTICA E CONSTRUÇÃO DE CONFIANÇA EM MODELOS DE INTELIGÊNCIA ARTIFICIAL

Lucas Cerqueira Machado Dias
Centro de Estudos e Sistemas Avançados do Recife

Gabriela de Araújo Macedo
Universidade de Pernambuco

Laura Galvão Marques Cavalcante
Centro Universitário Maurício de Nassau

RESUMO

Aborda a ética e a construção de confiança em modelos de inteligência artificial (IA), explorando desafios e oportunidades na construção de uma IA ética, confiável, transparente, justa e segura. O objetivo principal é analisar a interseção entre ética e construção de confiança em IA, discutindo questões como vieses, discriminação, privacidade, responsabilidade, impacto social e mecanismos para aumentar a confiabilidade dos sistemas de IA. Os métodos utilizados na pesquisa envolvem a revisão bibliográfica e documental, com análise de casos e exemplos para ilustrar os desafios e oportunidades na construção de uma IA ética e confiável. Os resultados da pesquisa apontam para a necessidade de uma abordagem holística para a construção de IA ética e confiável, que englobe a mitigação de vieses, a proteção da privacidade, a explicabilidade, a responsabilização e o diálogo com as partes interessadas. Destaca a importância da ética em todo o ciclo de vida da IA, desde a concepção até a implementação e o uso, e a necessidade de integrar princípios éticos no design e desenvolvimento de sistemas de IA. A pesquisa conclui que a construção de confiança em IA exige esforços conjuntos, incluindo a integração de princípios éticos no design dos sistemas, o desenvolvimento de mecanismos de explicabilidade e transparência, a promoção da diversidade e inclusão nas equipes de desenvolvimento, a criação de “regulatory sandboxes” para testar novas tecnologias e a implementação de regulamentações governamentais que promovam a IA responsável.

Palavras-chave: ética; confiança; inteligência artificial; vieses.

INTRODUÇÃO

A inteligência artificial (IA) está rapidamente se tornando uma força poderosa na sociedade moderna, impactando diversos setores como saúde, educação, justiça e mercado de trabalho. Com essa crescente influência, surgem questões éticas cruciais que exigem atenção e debate. O desenvolvimento e a implementação responsável da IA dependem da capacidade de garantir que os sistemas sejam éticos e confiáveis. Um conceito fundamental para alcançar esse objetivo é a construção de confiança, que se refere aos mecanismos e estratégias utilizados para gerar confiança nos sistemas de IA, assegurando que sejam transparentes, justos, seguros e responsáveis.

Este trabalho aborda a interseção entre ética e construção de confiança em modelos de IA, explorando os desafios e oportunidades na construção de uma IA que seja não apenas eficiente, mas também justa, transparente e segura. A ética em IA abrange questões como vieses, discriminação, privacidade, responsabilidade e impacto social, enquanto a construção de confiança se concentra em mecanismos para avaliar e aumentar a confiabilidade dos sistemas de IA, como métricas de avaliação, explicabilidade e interpretabilidade. Para iniciarmos essa discussão de forma mais impactante, podemos considerar o caso da ferramenta de recrutamento da Amazon, que foi abandonada por apresentar vieses de gênero ¹. Esse exemplo ilustra como a falta de atenção aos aspectos éticos pode levar a consequências negativas e prejudicar a confiança na IA.

MÉTODOS

Vieses e Discriminação em Modelos de IA

Um dos principais desafios éticos na IA é a presença de vieses e discriminação em modelos de IA. Vieses podem surgir de diversas fontes, incluindo dados de treinamento enviesados, design algorítmico inadequado e falta de diversidade nas equipes de desenvolvimento (MANYIKA; SILBERG; PRESTEN, 2022). É importante analisar a taxonomia dos vieses, que se dividem em vieses humanos e vieses algorítmicos, para entender como esses vieses influenciam a formação de generalizações, estereótipos e preconceitos (LUCENA, 2023).

Algoritmos de machine learning podem perpetuar e amplificar vieses existentes nos dados, levando a resultados discriminatórios em áreas como

recrutamento, justiça criminal e concessão de crédito (LUCENA, 2023). Por exemplo, um sistema de IA treinado com dados históricos que refletem desigualdades raciais pode perpetuar essas desigualdades ao tomar decisões sobre contratações ou sentenças judiciais (SOUSA; PINHEIRO, 2019).

Contudo, a IA também tem o potencial de reduzir a interpretação subjetiva de dados, tornando o processo decisório mais imparcial (MANYIKA; SILBERG; PRESTEN, 2022). Algoritmos de machine learning podem ser utilizados para identificar e mitigar vieses humanos, contribuindo para decisões mais justas e equitativas.

É crucial garantir que os modelos de IA sejam justos e imparciais, utilizando técnicas de mitigação de vieses, como a coleta de dados mais representativos, o desenvolvimento de algoritmos mais robustos e a auditoria regular dos sistemas em busca de vieses (GOMES, 2021). A revisão humana também desempenha um papel importante na garantia da justiça e da não discriminação em sistemas de IA (LUCENA, 2023).

Métricas e Métodos para Avaliar a Confiabilidade de Modelos de IA

A construção de confiança em IA depende da capacidade de avaliar a confiabilidade dos modelos. Métricas e métodos de avaliação são essenciais para determinar se um sistema de IA é preciso, confiável e seguro (SILVA, 2022).

Acurácia, precisão, sensibilidade e especificidade são algumas das métricas comumente utilizadas para avaliar o desempenho de modelos de IA (SILVA; FERREIRA, 2023). Além disso, é importante considerar a robustez do modelo, ou seja, sua capacidade de lidar com dados inesperados ou adversários, e sua estabilidade, ou seja, sua capacidade de manter o desempenho ao longo do tempo (DATACAMP, s.d.).

A qualidade dos dados utilizados para treinar e avaliar os modelos de IA é fundamental para garantir a confiabilidade dos resultados (ASTERA, s.d.). Métricas de qualidade de dados, como completude, consistência e precisão, devem ser utilizadas para avaliar a integridade dos dados e garantir que os modelos sejam treinados com dados confiáveis.

A explicabilidade e a interpretabilidade são cruciais para a construção de confiança em IA. A capacidade de entender como um modelo de IA chega a

uma determinada decisão é fundamental para garantir a confiança do usuário e permitir a auditoria e a responsabilização (MANAGEMENT SOLUTIONS, s.d.).

Modelos de caixa preta, como redes neurais profundas, podem ser difíceis de interpretar, o que dificulta a compreensão de seu funcionamento e a identificação de potenciais vieses (MANAGEMENT SOLUTIONS, s.d.). Técnicas de explicabilidade, como LIME e SHAP, buscam tornar os modelos mais transparentes, fornecendo insights sobre como as variáveis de entrada influenciam as previsões (CARVALHO, 2022).

Outras técnicas, como PDP (Partial Dependence Plots) e Anchors, também contribuem para a interpretabilidade dos modelos, fornecendo diferentes perspectivas sobre o processo decisório da IA (MANAGEMENT SOLUTIONS, s.d.).

É importante reconhecer que, apesar dos avanços na área de explicabilidade de IA (XAI), ainda existem desafios na interpretação de modelos complexos (MANAGEMENT SOLUTIONS, s.d.). A reprodutibilidade dos resultados, a presença de vieses nos dados de entrada e a precisão da explicação são fatores que podem limitar a interpretabilidade dos modelos.

A IA responsável, que busca garantir que a IA seja utilizada de forma ética e responsável, tem objetivos similares à XAI, mas com abordagens diferentes (IBM, s.d.). Enquanto a XAI se concentra em explicar os resultados da IA após serem computados, a IA responsável busca integrar princípios éticos no design e desenvolvimento dos sistemas, tornando-os responsáveis desde o início.

A calibração e a incerteza são aspectos importantes a serem considerados na avaliação de modelos de IA. A calibração se refere à capacidade do modelo de fornecer previsões probabilísticas precisas, enquanto a incerteza se refere à confiança do modelo em suas previsões (LOPES, 2020).

Modelos bem calibrados fornecem previsões probabilísticas que correspondem à frequência real dos eventos, enquanto modelos que expressam incerteza de forma adequada permitem aos usuários avaliar o risco associado às previsões (FARIA, s.d.). A calibração é crucial para garantir a confiabilidade das previsões, especialmente em aplicações que exigem alta precisão, como na área médica (RIBEIRO, 2017).

A incerteza em medição e calibração está relacionada à "dúvida" inerente a qualquer processo de medição (SOFTEXPERT, 2021). É fundamental considerar as fontes de incerteza, como o instrumento de medição, o operador,

as condições ambientais e o procedimento utilizado, para avaliar a confiabilidade das medições.

A calibração e a incerteza são especialmente importantes em aplicações de alto risco, como diagnósticos médicos e veículos autônomos (LOPES, 2020). Nesses casos, é crucial que o modelo seja capaz de fornecer previsões precisas e expressar a incerteza de forma adequada, para que os usuários possam tomar decisões informadas e seguras.

RESULTADOS

A crescente utilização de dados em sistemas de IA levanta preocupações sobre privacidade e proteção de dados. A coleta, armazenamento e processamento de grandes volumes de dados, incluindo informações pessoais sensíveis, exigem medidas rigorosas para garantir a privacidade e segurança dos indivíduos (POETA, 2019).

A Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na Europa estabelecem diretrizes para o tratamento de dados pessoais, incluindo a necessidade de consentimento, transparência e medidas de segurança (RODRIGUES, 2021). No contexto da IA, é fundamental garantir que os sistemas estejam em conformidade com as leis de proteção de dados e que os dados sejam utilizados de forma responsável e ética.

Um conceito importante nesse contexto é o "Privacy by Design", que defende a incorporação da privacidade desde as etapas iniciais de design de sistemas de IA (MENDES; SANTOS, 2022). Isso significa considerar a privacidade como um requisito fundamental e implementar medidas de proteção de dados ao longo de todo o ciclo de vida do sistema.

Outro ponto relevante é a "decisão automatizada", que se refere a decisões tomadas por sistemas de IA sem intervenção humana. A LGPD aborda essa questão, estabelecendo direitos e garantias para os indivíduos em relação a decisões automatizadas que afetem seus interesses (RODRIGUES, 2021).

A IA generativa, que se refere a modelos de IA capazes de gerar novos dados, como imagens, textos e vídeos, apresenta desafios específicos para a privacidade e proteção de dados (MENDES; SANTOS, 2022). É crucial garantir que os dados utilizados para treinar esses modelos sejam coletados e utilizados

de forma ética e responsável, e que os modelos não sejam utilizados para gerar conteúdo falso ou enganoso.

Responsabilidade e Prestação de Contas em IA

A questão da responsabilidade e prestação de contas em IA é complexa e desafiadora. A medida que os sistemas de IA se tornam mais autônomos e complexos, torna-se crucial determinar quem é responsável pelas decisões tomadas por esses sistemas (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA, 2021).

A atribuição de responsabilidade em casos de danos ou erros causados por sistemas de IA exige uma análise cuidadosa do ciclo de vida da IA, desde o design e desenvolvimento até a implementação e uso (FERREIRA; SILVA, 2024). É necessário estabelecer mecanismos de responsabilização para garantir que as partes interessadas, incluindo desenvolvedores, empresas e usuários, sejam responsáveis pelas consequências das ações da IA.

Um dos desafios na atribuição de responsabilidade é a dificuldade de prever e controlar o comportamento de sistemas de IA complexos (UNICEP, 2023). A opacidade de alguns modelos de IA, como as redes neurais profundas, pode dificultar a compreensão de como o sistema chegou a uma determinada decisão, o que torna a atribuição de culpa mais complexa.

No contexto da IA, a responsabilidade se refere ao papel das pessoas e à capacidade dos sistemas de IA de responder por uma decisão e identificar erros ou resultados inesperados, enquanto a prestação de contas se refere à necessidade do sistema de IA explicar e justificar suas decisões e ações para seus parceiros, usuários e outros com quem o sistema interage (BORTOLOZZI; SILVA, 2023).

É fundamental desenvolver mecanismos de responsabilização que levem em consideração a complexidade da IA e garantam a justiça e a equidade na atribuição de culpa. A proteção de denunciadores também é crucial para garantir a responsabilização e a ética no desenvolvimento e implementação de sistemas de IA (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA, 2021).

Aumentar a confiança do usuário em sistemas de IA é essencial para a adoção e o uso responsável da IA. Diversos mecanismos podem contribuir para a construção de confiança, incluindo:

- **Transparência:** Fornecer informações claras e acessíveis sobre o funcionamento do sistema, os dados utilizados e as decisões tomadas. Isso aumenta a compreensão do usuário sobre o sistema, permitindo que ele avalie sua confiabilidade e seus potenciais riscos.
- **Explicabilidade:** Permitir que os usuários compreendam como o sistema de IA chega a uma determinada decisão. Facilita a identificação de vieses e erros, aumenta a confiança nas decisões e permite a auditoria do sistema.
- **Controle do usuário:** Dar aos usuários algum grau de controle sobre o sistema de IA, permitindo que personalizem o comportamento do sistema ou intervenham em suas decisões. Aumenta a sensação de controle e autonomia do usuário, tornando o sistema mais aceitável e confiável.
- **Segurança:** Garantir que o sistema de IA seja seguro e resistente a ataques, protegendo os dados do usuário e evitando consequências negativas. Protege a privacidade e a segurança do usuário, evitando o uso indevido do sistema e suas informações.
- **Responsabilidade:** Estabelecer mecanismos de responsabilização para garantir que as partes envolvidas no desenvolvimento e uso da IA sejam responsáveis por suas ações. Promove a ética e a responsabilidade no desenvolvimento e uso da IA.
- **Design centrado no humano:** Desenvolver sistemas de IA que atendam às necessidades, valores e experiências dos usuários. Isso aumenta a usabilidade, a aceitabilidade e a confiança nos sistemas de IA, tornando-os mais eficazes e benéficos para os usuários.
- **Monitoramento contínuo:** Analisar logs e monitorar métricas para garantir o desempenho do sistema e identificar potenciais problemas. Permite detectar e corrigir vieses, melhorar o desempenho do sistema e manter a confiança do usuário ao longo do tempo.
- **Controlabilidade:** Implementar mecanismos para monitorar e orientar o comportamento do sistema de IA. Assegura que o sistema opere dentro

de limites aceitáveis, evitando consequências negativas e permitindo a intervenção humana quando necessário.

DISCUSSÃO

A IA está transformando o mercado de trabalho, automatizando tarefas e criando novas oportunidades, mas também gerando preocupações sobre o deslocamento de trabalhadores e o aumento da desigualdade (BANDEIRA, 2020). A IA tem o potencial de aumentar a produtividade e a eficiência em diversos setores, mas também pode levar à substituição de mão de obra humana por máquinas (CUNHA; FONSECA, 2021).

É importante considerar o impacto social da IA e desenvolver estratégias para mitigar os efeitos negativos, como programas de requalificação profissional e políticas de apoio aos trabalhadores afetados pela automação (CUNHA; FONSECA, 2021). Além disso, é crucial garantir que a IA seja utilizada para promover o bem-estar social e reduzir as desigualdades, em vez de exacerbá-las.

O progresso tecnológico que a IA possibilita apresenta desafios historicamente sem precedentes para as empresas e, em especial, para todos os trabalhadores (CARVALHO, 2018). A IA tem o potencial de redefinir as relações entre empregadores e trabalhadores, impactando a dinâmica socioeconômica (SILVA, 2019).

A tributação internacional de empresas de IA também apresenta desafios, pois a natureza globalizada dessas empresas dificulta a aplicação das leis tributárias tradicionais (BANDEIRA, 2020). É necessário criar mecanismos de tributação justos e eficientes que garantam que as empresas de IA contribuam para o desenvolvimento social e econômico dos países onde operam.

A IA tem o potencial de revolucionar a área da saúde, com aplicações que vão desde o diagnóstico médico até o desenvolvimento de novos tratamentos. A IA pode ser utilizada para analisar grandes volumes de dados médicos, como imagens de exames e registros de pacientes, para auxiliar no diagnóstico de doenças, prever riscos à saúde e personalizar tratamentos (BORTOLZZI; SILVA, 2023).

A IA pode transformar a educação, personalizando o aprendizado, automatizando tarefas e fornecendo novas oportunidades educacionais. Sistemas de tutoria inteligente podem adaptar o conteúdo e o ritmo de aprendizado

às necessidades individuais dos alunos, enquanto plataformas de ensino à distância podem utilizar a IA para fornecer feedback personalizado e avaliar o progresso dos alunos.

A IA pode ser uma ferramenta poderosa para enfrentar os desafios ambientais, com aplicações em áreas como mitigação das mudanças climáticas, gestão de recursos e redução da poluição. A IA pode ser utilizada para monitorar o desmatamento, otimizar o consumo de energia, desenvolver tecnologias de energia renovável e prever desastres naturais.

A IA pode desempenhar um papel crucial na cibersegurança, ajudando a proteger sistemas e dados contra ameaças cibernéticas. Algoritmos de IA podem ser utilizados para detectar atividades suspeitas, identificar vulnerabilidades e responder a incidentes de segurança de forma mais rápida e eficiente.

A construção de IA ética e confiável exige uma abordagem holística que engloba os diversos aspectos discutidos neste relatório. É fundamental integrar princípios éticos no design e desenvolvimento de sistemas de IA, utilizando técnicas de mitigação de vieses, garantindo a privacidade e segurança dos dados, promovendo a explicabilidade e a interpretabilidade, e estabelecendo mecanismos de responsabilização (SERPRO, 2023)

A ética deve estar presente em todo o ciclo de vida da IA, desde a concepção até a implementação e o uso, seguindo o princípio da “ética by design” (SERPRO, 2023) Isso significa considerar as implicações éticas em todas as etapas do desenvolvimento da IA, garantindo que os sistemas sejam projetados para serem justos, transparentes e responsáveis.

Além disso, é crucial envolver as partes interessadas, incluindo especialistas em ética, cientistas sociais, usuários e o público em geral, no debate sobre a IA ética, promovendo a transparência e a participação na tomada de decisões (GARCIA; MARIATH, 2022). A IA ética e confiável é um objetivo que exige colaboração e diálogo entre diferentes áreas do conhecimento e setores da sociedade.

Políticas de transparência, especialmente em relação a recursos e interações que envolvem o poder público, também podem contribuir para a consolidação de uma postura ética na IA (LEITE, 2023). É importante que as organizações sejam transparentes sobre como os sistemas de IA são utilizados, quais dados são coletados e como as decisões são tomadas.

A diversidade e a inclusão nas equipes de desenvolvimento de IA são essenciais para garantir que os sistemas sejam justos e representem a diversidade da sociedade (PROGRAMA ÉTICA, s.d.). Equipes diversas podem trazer diferentes perspectivas e experiências, ajudando a identificar e mitigar vieses e a construir sistemas mais inclusivos.

A construção de confiança desempenha um papel fundamental no desenvolvimento de IA ética ao fornecer ferramentas e métodos para avaliar e aumentar a confiabilidade dos sistemas de IA. Ao garantir que os sistemas sejam transparentes, explicáveis, justos e seguros, a construção de confiança contribui para a construção de uma IA que seja não apenas eficiente, mas também ética e responsável (ALMEIDA, 2024).

A construção de confiança pode ajudar a identificar e mitigar vieses, proteger a privacidade dos dados, aumentar a confiança do usuário e promover a responsabilização (SOUSA; PINHEIRO, 2022). Ao fornecer uma base sólida para a avaliação e o controle dos sistemas de IA, a construção de confiança contribui para o desenvolvimento de uma IA que esteja alinhada com os valores éticos e sociais.

Um mecanismo inovador para promover a construção de confiança e o desenvolvimento de IA ética são os “regulatory sandboxes” (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA, 2021). Essas “caixas de areia regulatórias” permitem que empresas e pesquisadores testem novas tecnologias de IA em um ambiente controlado, com acompanhamento e supervisão de órgãos reguladores, o que ajuda a identificar e mitigar riscos e a promover a inovação responsável.

CONCLUSÃO

A ética e a construção de confiança são elementos essenciais para o desenvolvimento e a implementação responsável da IA. A construção de sistemas de IA que sejam éticos e confiáveis exige uma abordagem holística que englobe a mitigação de vieses, a proteção da privacidade, a explicabilidade, a responsabilização e o diálogo com as partes interessadas.

A IA ética e confiável oferece um futuro promissor para a sociedade, com o potencial de solucionar problemas complexos, melhorar a qualidade de

vida e promover o bem-estar social. No entanto, é crucial estarmos atentos aos desafios éticos, como a complexidade dos sistemas, a necessidade de equilibrar inovação com segurança e o potencial de exacerbar desigualdades sociais.

A construção de confiança em IA exige uma combinação de esforços, incluindo a integração de princípios éticos no design dos sistemas ("ética by design"), o desenvolvimento de mecanismos de explicabilidade e transparência, a promoção da diversidade e inclusão nas equipes de desenvolvimento, a criação de "regulatory sandboxes" para testar novas tecnologias e a implementação de regulamentações governamentais que promovam a IA responsável.

A colaboração entre diferentes áreas do conhecimento, como ética, direito, ciência da computação e ciências sociais, é fundamental para garantir que a IA seja desenvolvida e utilizada de forma ética, responsável e benéfica para todos. A IA ética e confiável é um objetivo que exige um esforço conjunto da sociedade para garantir que essa poderosa tecnologia seja utilizada para promover o progresso humano e a justiça social.

Agradecimentos

Agradeço profundamente à minha família, meu alicerce inabalável, pelo apoio incondicional durante todo o desenvolvimento desta pesquisa. Em especial, à minha amada esposa Laura, cuja dedicação, paciência e incentivo foram essenciais para que eu superasse os desafios e concluísse esta jornada. Seu amor e compreensão foram a força motriz que me impulsionou a cada passo.

REFERÊNCIAS

Astera. (s.d.). **7 métricas de qualidade de dados para avaliar a integridade dos seus dados**. Recuperado em 22 de dezembro de 2024, de <https://www.astera.com/pt/type/blog/data-quality-metrics/>

Alura. (2023, 21 de março). **Ética e inteligência artificial: qual a relação e os desafios?** <https://www.alura.com.br/artigos/etica-e-inteligencia-artificial>

Almeida, V. (2024). Inteligência artificial e o futuro do trabalho. **Jornal da USP**. <https://jornal.usp.br/wp-content/uploads/2024/05/1-Virgilio-Almeida.pdf>

Amazon Web Services. (2023, 18 de maio). Melhores práticas de segurança em projetos de IA generativa. **Blog da AWS**. <https://aws.amazon.com/pt/blogs/aws-brasil/melhores-praticas-de-seguranca-em-projetos-de-ia-generativa/>

Bandiera, L. C. J. F. (2020). A proteção de dados pessoais no contexto da internet das coisas: desafios e perspectivas [Dissertação de Mestrado, Universidade do Vale do Itajaí]. Repositório Institucional do Univali. <https://www.univali.br/Lists/TrabalhosMestrado/Attachments/3223/Disserta%C3%A7%C3%A3o%20-%20Lucas%20Cezar%20Jos%C3%A9%20Figueiredo%20Bandiera.pdf>

Bortolozzi, F., & Silva, R. W. da. (2023). Inteligência artificial e sociedade: avanços e riscos. **Estudos Avançados**, 37(107). <https://www.scielo.br/j/ea/a/c4sqqrthGMS3ngdBhGWtKhH/>

Bortolozzi, F., & Silva, R. W. da. (2023). Inteligência artificial: riscos, benefícios e uso. **Estudos Avançados**, 37(108). <https://www.scielo.br/j/ea/a/ZnKyrCrLVqzhZbXGgXTwDtn>

Carvalho, A. P. de. (2018). Percepção do impacto da inteligência artificial em contexto ocupacional impact of artificial intelligence on the workplace [Dissertação de Mestrado, Universidade de Lisboa]. Repositório Comum. <https://comum.rcaap.pt/bitstream/10400.26/29226/1/2018%20Impact%20of%20artificial%20intelligence%20on%20the%20workplace.pdf>

Carvalho, N. D. de. (2022). MoReXAI: um modelo para refletir sobre inteligência artificial explicável [Dissertação de Mestrado, Universidade Federal do Ceará]. Repositório Institucional UFC. https://repositorio.ufc.br/bitstream/riufc/70423/1/2022_dissertacao_ndeocarvalho.pdf

Comissão Europeia. (2020). Livro branco sobre a inteligência artificial. Uma abordagem europeia virada para a excelência e a confiança. https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_pt?filename=commission-white-paper-artificial-intelligence-feb2020_pt.pdf

Cruz, M. da C. (2023). **A transparência no centro da construção de uma ia ética**. *Novos Estudos CEBRAP*, (126). <https://www.scielo.br/j/nec/a/pD9k5gtHpXwsgFcsMC56jBg/>

Cunha, M. A. da, & Fonseca, T. da. (2021). Inteligência artificial, trabalho e produtividade. **Revista de Administração de Empresas**, 61(3). <https://www.scielo.br/j/rae/a/WcDHgCMrtR5RyqYVX6gGZkQ/>

DataCamp. (s.d.). **Avaliação do LLM: Métricas, metodologias, práticas recomendadas**. Recuperado em 22 de dezembro de 2024, de <https://www.datacamp.com/pt/blog/llm-evaluation>

Escola Nacional de Administração Pública. (2022). Regulação da inteligência artificial. **Repositório Institucional da ENAP**. <https://repositorio.enap.gov.br/bitstream/1/7419/1/2022.12.08%20-%20Regula%C3%A7%C3%A3o%20da%20Intelig%C3%Aancia%20Artificial.pdf>

Faria, S. B. de. (s.d.). **Tratamento de incerteza**. Universidade Federal de Santa Catarina. http://www.inf.ufsc.br/~silvia/disciplinas/sep/material_didatico/MaterialDidatico.pdf

Ferreira, S. C. S., & Silva, F. C. M. da. (2024). **Por um uso ético da inteligência artificial em periódicos científicos**. BVS SES-SP. <https://ses.sp.bvs.br/periodicosp/wp-content/uploads/2024/08/Usa-Etico-da-Inteligencia-Artificial-em-Periodicos-Cientificos.pptx-2.pdf>

Garcia, L. F. M., & Mariath, R. C. (2022). **Regulação de inteligência artificial no Brasil**. FGV Direito Rio. <https://direitorio.fgv.br/sites/default/files/2022-08/policypaperiaegoverno.pdf>

Gomes, A. R. (2021). Inteligência artificial e a discriminação algorítmica: uma análise do caso Amazon. **Revista Eletrônica Direito & TI**, (12). <https://direitoeti.com.br/direitoeti/article/view/212>

IBM. (s.d.). Exemplos de vieses de IA. Recuperado em 22 de dezembro de 2024, de <https://www.ibm.com/br-pt/think/topics/shedding-light-on-ai-bias-with-real-world-examples>

IBM. (s.d.). **O que é IA explicável (XAI)?**. Recuperado em 22 de dezembro de 2024, de <https://www.ibm.com/br-pt/topics/explainable-ai>

Lamb, L. (2024). Ética em IA e IA ética: prolegômenos e estudo de casos significativos. **Jornal da USP**. <https://jornal.usp.br/wp-content/uploads/2024/05/8-Luis-Lamb-certo.pdf>

Leite, J. P. (2023). O uso da inteligência artificial no compliance anticorrupção corporativo [Dissertação de Mestrado, Fundação Getulio Vargas]. eduCAPES. <https://educapes.capes.gov.br/bitstream/capes/744825/1/o-uso-da-inteligencia-artificial-no-compliance-anticorrupcao-corporativo.pdf>

Lopes, H. (2020, 10 de agosto). Quantificando incerteza de modelos de regressão com previsão conforme. **Medium**. <https://medium.com/data-hackers/quantificando-incerteza-de-modelos-de-regress%C3%A3o-com-previs%C3%A3o-conforme-766fa03b2587>

Lucena, A. M. (2023). **Discriminação algorítmica**: inteligência artificial, vieses humanos e algorítmicos e a proteção constitucional algorithmic d. *Direito em Debate*, 42(87). <https://online.unisc.br/seer/index.php/direito/article/view/18961/11493>

Management Solutions. (s.d.). Explainable artificial intelligence (XAI) - Desafios na interpretabilidade de modelos. Recuperado em 22 de dezembro de 2024, de <https://www.managementsolutions.com/sites/default/files/minisite/static/22959b0f-b3da-47c8-9d5c-80ec3216552b/iax/pdf/explainable-artificial-intelligence-pt.pdf>

Management Solutions. (s.d.). **Técnicas de interpretabilidade**: estado da arte - Explainable artificial intelligence (XAI). Recuperado em 22 de dezembro de 2024, de <https://www.managementsolutions.com/sites/default/files/minisite/static/22959b0f-b3da-47c8-9d5c-80ec3216552b/iax/pdf/explainable-artificial-intelligence-pt-04.pdf>

Manyika, J., Silberg, J., & Presten, B. (2022, 14 de novembro). **Como lidar com vieses na inteligência artificial (e nos seres humanos)**. McKinsey & Company. <https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans/pt-BR>

Marques, A. L. (2021). O direito fundamental à explicabilidade da inteligência artificial utilizada em decisões estatais. **Revista da AGU**, (118). <https://revistaagu.agu.gov.br/index.php/AGU/article/view/3480/2934>

Mendes, M. I. S., & Santos, B. P. dos. (2022). Inteligência artificial e privacidade: os desafios do privacy by design. **Advances in Knowledge Representation**, 3(1). <https://periodicos.ufmg.br/index.php/advances-kr/article/download/52813/44775/199635>

Oliveira, L. de. (2020). O impacto da inteligência artificial no mundo do trabalho. **Revista Multidisciplinar do Nordeste Mineiro**, 7(3). <https://revista.unipacto.com.br/index.php/multidisciplinar/article/view/1682>

Organização das Nações Unidas para a Educação, a Ciência e a Cultura. (2021). Recomendação sobre a ética da inteligência artificial. <https://idl-bnc-idrc.dspacedirect.org/bitstreams/4c4bed79-4c84-46e0-bec6-aacf04491465/download>

Organização das Nações Unidas para a Educação, a Ciência e a Cultura. (2021). Recomendação sobre a ética da inteligência artificial. UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000381137_por

Poeta, V. S. (2019). Reflexos do Regulamento Geral de Proteção de Dados Europeu (RGPD) na proteção de dados pessoais de crianças e adolescentes [Dissertação de Mestrado, Universidade do Vale do Itajaí]. Repositório Institucional da Univali. <https://www.univali.br/Lists/TrabalhosMestrado/Attachments/3015/DISSERTA%C3%87%C3%83O%20-%20VITOR%20SARDAGNA%20POETA.pdf>

Programa Ética. (s.d.). Ética na inteligência artificial: desafios e oportunidades. Recuperado em 22 de dezembro de 2024, de <https://programae.org.br/imprensa/etica-na-inteligencia-artificial-desafios-e-oportunidades/>

Ribeiro, C. (2017). Cálculo e análise de incerteza associada à calibração de instrumentos de medição instalados em linha [Dissertação de Mestrado, Instituto Politécnico de Leiria]. Repositório IPL. https://repositorio.ipl.pt/bitstream/10400.21/165voc/1/C%C3%A9sar%20Ribeiro_43289_DEQ.pdf

Rodrigues, N. (2021). Inteligência artificial e lei geral de proteção de dados pessoais: o direito à explicação nas decisões automatizadas. **Revista Brasileira de Direito Civil**, 17(3), 425–454. <https://rbdcivil.ibdcivil.org.br/rbdc/article/download/584/425/1992>

Serpro. (2023, 27 de junho). Ética na IA. <https://www.serpro.gov.br/menu/noticias/noticias-2023/etica-na-ia>

Silva, A. F. da. (2019). O impacto da inteligência artificial no mercado de trabalho [Dissertação de Mestrado, Pontifícia Universidade Católica de Goiás]. Repositório Institucional da PUC Goiás. <https://repositorio.pucgoias.edu.br/jspui/handle/123456789/8153>

Silva, F. C. M. da, & Ferreira, S. C. S. (2023). Habilidades informacionais necessárias ao uso de ferramentas de ia generativa em demandas informacionais de natureza acadêmica-científica inteligência artificial (ia) generativa e competência em informação. **Perspectivas em Ciência da Informação**, 28(3). <https://www.scielo.br/j/pci/a/GVCW7KbcRjGVhLSrmy3PCng/>

Silva, J. M. da. (2022). Métricas utilizadas para avaliar a eficiência de classificadores em algoritmos inteligentes [Trabalho de Conclusão de Curso, Universidade Federal do Rio Grande do Norte]. ResearchGate. https://www.researchgate.net/publication/359541310_METRICAS_UTILIZADAS_PARA_AVALIAR_A_EFICIENCIA_DE_CLASSIFICADORES_EM_ALGORITMOS_INTELIGENTES

SoftExpert. (2021, 17 de fevereiro). O que é incerteza em medição e calibração. **Blog SoftExpert**. <https://blog.softexpert.com/pt-br/incerteza-em-medicao-calibracao/>

Sousa, R. M. de, & Pinheiro, P. G. (2019). A ética na inteligência artificial: desafios ethics of artificial intelligence: challenges. **International Journal of Advanced Engineering Research and Science**, 6(11). https://www.researchgate.net/publication/336680544_A_Etica_na_Inteligencia_Artificial_Desafios_Ethics_of_Artificial_Intelligence_Challenges

Sousa, R. M. de, & Pinheiro, P. G. (2022). O uso da inteligência artificial como ferramenta de eficiência e acesso à justiça em revisão sistemática da literatura. **Revista de Direito Brasileira**, 19(3). https://www.researchgate.net/publication/363033739_O_uso_da_inteligencia_artificial_como_ferramenta_de_eficiencia_e_acesso_a_justica_em_revisao_sistematica_da_literatura

Unicep. (2023, 17 de outubro). Inteligência artificial e ética: qual deve ser o impacto ético para os próximos anos? <https://www.unicep.edu.br/post/inteligencia-artificial-e-etica-qual-deve-ser-o-impacto-etico-para-os-proximos-anos>

Unite.AI. (s.d.). Como a IA explicável aumenta a confiabilidade e a confiabilidade. Recuperado em 22 de dezembro de 2024, de <https://www.unite.ai/pt/como-a-IA-explic%C3%A1vel-aumenta-a-confiabilidade-e-a-confiabilidade/>

ÉTICA E INTELIGÊNCIA ARTIFICIAL: OS DESAFIOS DO PRINCÍPIO DA EXPLICABILIDADE À LUZ DO PL 2.338/23

Gabriela de Araújo Macedo
Universidade de Pernambuco (UPE)

Lucas Cerqueira Machado Dias
Centro de Estudos e Sistemas Avançados do Recife

Laura Galvão Marques Cavalcante
Centro Universitário Maurício de Nassau

RESUMO

O presente artigo busca analisar a forma como os sistemas de inteligência artificial (IA) se alinham aos princípios éticos delineados no Projeto-lei nº 2.338/23, doravante chamado de PL 2.338/23, com foco específico nos desafios tecnológicos em atender ao princípio da explicabilidade. O objetivo dessa abordagem será identificar na normativa os requisitos necessários para que o princípio em comento seja devidamente atendido, bem como explorar os conceitos, técnicas e desafios que envolvem o desenvolvimento de sistemas explicáveis de inteligência artificial. A metodologia adotada foi a pesquisa documental em torno dos dispositivos legais que abordam e regulamentam o princípio ético em análise, bem como a revisão bibliográfica em busca de analisar os conceitos que permeiam as áreas da IA exploradas no presente trabalho, investigando também possíveis caminhos para dirimir os desafios tecnológicos encontrados na construção desses tipos de sistemas.

Palavras-chave: ética; inteligência artificial; explicabilidade; projeto-lei nº 2.338/23.

INTRODUÇÃO

A inteligência artificial (IA) é um ramo da ciência da computação que abrange o desenvolvimento de sistemas inteligentes, podendo ser definida como um sistema de informação com capacidade de fazer previsões, recomendações e/ou decisões que influenciam ambientes reais ou virtuais para um determinado conjunto de objetivos previamente definidos por humanos (OECD, 2019). Consoante McCarthy *et al*, pode ser definida como a ciência de fazer máquinas inteligentes, especialmente programas de computador inteligentes, estando relacionado à tarefa de usar computadores para entender a inteligência humana, sem contudo precisar se limitar a métodos que são biologicamente observáveis. (McCarthy *et al*, 2006)

Os sistemas de IA vêm se consolidando no cenário mundial como uma das principais tecnologias disruptivas da atualidade, influenciando desde áreas como marketing e publicidade até cenários de alto risco como intervenções na saúde, sendo amplamente utilizada no auxílio de diagnósticos e procedimentos cirúrgicos. À vista de papel tão fundamental na organização social contemporânea, diversos países têm se empenhado para elaborar instruções normativas que visem alinhar o uso da inteligência artificial com os princípios éticos arraigados na sociedade a fim de mitigar seus impactos negativos e dirimir decisões incorretas, uma vez que esses sistemas se apoiam amplamente no tratamento de dados de pessoas naturais, podendo levar a vieses incorretos ou conclusões controversas.

Nessa toada, a UNESCO aprovou em novembro de 2021 a Recomendação sobre a Ética da Inteligência Artificial, considerada um marco global para orientar o uso e desenvolvimento ético desse tipo de sistema. A transparência aparece como um dos pilares essenciais dessa recomendação, enfatizando que os sistemas de IA devem ser desenvolvidos com base na divulgação de informações, explicabilidade e livre acesso a dados, vislumbrando o uso responsável da inteligência artificial de maneira que seja uma ferramenta para promover o bem estar social, dado que sistemas transparentes auxiliam na confiabilidade e aceitação por parte dos usuários. (UNESCO, 2021)

O primeiro marco regulatório a tratar sobre o tema surgiu na União Europeia com o chamado "AI Act", estabelecendo disposições normativas que, assim

como a recomendação mencionada, também busca alinhar o desenvolvimento e uso de sistemas de IA com os princípios éticos vigentes. O referido documento busca essencialmente garantir segurança, transparência, rastreabilidade, sustentabilidade e não discriminação nos sistemas de IA, elencando os requisitos de transparência necessários para que a IA esteja adequada aos padrões de ética social. Ainda que aprovado desde 2021, o “AI Act” apenas entrou em vigor recentemente, no mês agosto do presente ano. (UNIÃO EUROPEIA, 2024)

No cenário brasileiro, essa discussão ganhou força no ano de 2021 com a publicação da Estratégia Brasileira de Inteligência Artificial (EBIA). A EBIA elenca diretrizes a fim de direcionar o desenvolvimento desse tipo de sistema enquanto preserva preceitos éticos, buscando criar diretrizes que assegurem a transparência, segurança e privacidade no uso de IA. Ainda que trate de tópicos indispensáveis ao uso da IA, o documento tem função primordial de guia norteador para as empresas, não possuindo força normativa para regulamentar de fato o uso da IA no país. (BRASIL, 2021)

No campo jurídico, desde o ano de 2019 houveram propostas legislativas que tentaram dar início à regulamentação do tema, com destaque para o Projeto de Lei (PL) nº 5.052, de 2019, que buscava estabelecer os princípios para o uso da Inteligência Artificial no Brasil, o PL nº 21 de 2020, que visava estabelecer fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil, e por último, o PL nº 872, de 2021, que dispõe sobre o uso da Inteligência Artificial. (SENADO FEDERAL, p. 28, 2022)

Em vista de se tratarem de matérias em comum, os três referidos projetos passaram a tramitar conjuntamente no Senado Federal e, posteriormente, foi designada Comissão de Juristas visando a elaboração de texto legal mais avançado em substituição aos PLs mencionados, o qual foi apresentado somente em dezembro de 2022, surgindo o PL 2.338/23, objeto de análise do presente trabalho, que visa alinhar princípios, diretrizes e responsabilidades para o desenvolvimento e aplicação de sistemas de IA no Brasil. (SENADO FEDERAL, p. 29, 2022)

A respeito do PL 2.338/23, esse projeto traz como um de seus objetivos alinhar as estratégias de IA no país com as melhores práticas nacionais e internacionais, adotando um enfoque especial na busca por uma IA ética e responsável. A normativa em comento tem como cerne fomentar o desenvolvimento

da tecnologia traçando seus limites frente aos direitos fundamentais e sociais, determinando que o desenvolvimento tecnológico deve seguir as balizas já insculpidas no ordenamento jurídico e na sociedade, de maneira que as novas formas de organização social surgidas com a adoção dos sistemas de IA busquem pela primazia da ética e dos direitos fundamentais. (WACHOWICZ, 2024)

Nesse sentido, dentre os princípios trazidos pelo PL em comento, destaca-se aqui o princípio da explicabilidade a ser analisado no presente trabalho, o qual surge como um pilar fundamental dos objetivos éticos e de transparência trazidos na legislação. Importa ressaltar que a transparência e a publicidade já são princípios consagrados pela Constituição Federal como valores irrevogáveis do Estado Democrático de Direito, de modo que a lei apenas corrobora com essa máxima trazendo sua aplicação ao contexto de inovação tecnológica.

Nessa toada, o enfoque ora pesquisado será investigar os conceitos que permeiam a explicabilidade dos sistemas de inteligência artificial insculpada como princípio no PL 2.338/23, bem como analisar os desafios técnicos e tecnológicos para que essa premissa seja atendida. Visa-se analisar os dispositivos legais que se relacionam com a transparência e explicabilidade dos modelos de IA, investigando os requisitos a serem atendidos para o desenvolvimento ético dos sistemas, seguindo com o estudo do viés tecnológico por trás da compreensão dos sistemas de IA, apontando os possíveis desafios para que seja atingida. Discute-se também as vantagens aos usuários da implementação dessa abordagem, com um leve enfoque no setor público, para, ao final, analisar as possíveis estratégias a serem adotadas para mitigar o problema.

A pesquisa justifica-se pelo crescimento exponencial no uso e desenvolvimento dos sistemas de inteligência artificial no Brasil sem que haja regulamentação legal sobre a matéria. Em julho de 2024, o governo brasileiro anunciou um plano de investimento de 23 bilhões de reais para o setor de inteligência artificial, objetivando desenvolver tecnologias sustentáveis e de cunho social entre 2024 e 2028, demonstrando a relevância do tema e o impacto que a adoção desse tipo de sistema terá na organização social. (MINISTÉRIO DA CIÊNCIA TECNOLOGIA E INOVAÇÃO, 2024)

Portanto, afigura-se extremamente pertinente a presente análise acerca dos aspectos que permeiam o instituto da explicabilidade trazido pelo PL 2.338/23, na medida em que, caso aprovado, o Projeto-lei apresentará diversas consequências

ao uso e desenvolvimento desse tipo de sistema, contribuindo com a chamada IA responsável e alinhando a regulamentação no Brasil aos preceitos éticos já presentes em legislações ao redor do mundo.

MÉTODOS

Pesquisa documental

A pesquisa documental do presente estudo centrou-se em torno da análise do Projeto de Lei 2.338/23, buscando identificar de que maneira o princípio da explicabilidade é tratado no contexto da regulamentação da inteligência artificial no Brasil. Para tanto, foram examinados os dispositivos da normativa que abordam a transparência, a prestação de contas e a compreensibilidade das decisões automatizadas, a fim de avaliar se o texto normativo incorpora diretrizes alinhadas às exigências éticas de transparência mundialmente adotadas, de modo que, a fim de uma investigação mais robusta, a pesquisa considerou não apenas a literalidade do texto legal mas também sua razão de existir expressada pelo legislador.

Revisão bibliográfica

No presente estudo, a metodologia da revisão bibliográfica objetivou analisar a literatura existente no que tange aos conceitos da explicabilidade, na medida em que é o conceito cerne da análise do presente artigo. Para tal intuito, foram analisados artigos científicos, monografias e relatórios técnicos que discutem os fundamentos da Inteligência Artificial Explicável (xAI), suas aplicações e principalmente os desafios éticos envolvidos na transparência dos algoritmos. A seleção das fontes priorizou pesquisas recentes e relevantes na área, tendo em vista a crescente preocupação com a necessidade de sistemas de IA que possam fornecer justificativas compreensíveis a suas decisões.

Nessa toada, a revisão bibliográfica teve também finalidade de fornecer embasamento teórico à análise do Projeto de Lei 2.338/23, permitindo identificar se e como ele contempla o princípio em análise. Foram investigados estudos que analisam a regulamentação legislativa da inteligência artificial, discutindo a viabilidade técnica e os desafios práticos para implementação de mecanismos que visem uma maior transparência algorítmica. Desse modo, a

revisão permitiu uma fundamentação sólida acerca da explicabilidade como um requisito essencial para o desenvolvimento e a regulação da inteligência artificial de maneira ética.

RESULTADOS

A explicabilidade à luz do PL 2.338/23

Por explicabilidade define-se que todos os dados que conduziram à decisão automatizada, assim como o próprio modelo automático, podem ser fornecidos aos usuários finais e outras partes interessadas em termos não técnicos (MACOHIN, 2021). Essa compreensão é essencial para garantir que os cidadãos, empresas e governos possam compreender e confiar nas tecnologias que suportam decisões vitais. Sob a ótica do PL 2.883/23, a transparência não é apenas uma boa prática, mas uma obrigação legal e ética, indispensável para mitigar riscos como discriminação algorítmica, manipulação de informações e violações de privacidade.

O objetivo do PL é mencionado pelo relatório da seguinte forma (SENADO FEDERAL, 2022, p. 10):

Nessa quadra, esse novo marco legal tem um duplo objetivo. De um lado, estabelecer direitos para proteção do elo mais vulnerável em questão, a pessoa natural que já é diariamente impactada por sistemas de inteligência artificial, desde a recomendação de conteúdo e direcionamento de publicidade na Internet até a sua análise de elegibilidade para tomada de crédito e para determinadas políticas públicas. De outro lado, ao dispor de ferramentas de governança e de um arranjo institucional de fiscalização e supervisão, criar condições de previsibilidade acerca da sua interpretação e, em última análise, segurança jurídica para inovação e o desenvolvimento econômico-tecnológico.

Nesse sentido, o Projeto-lei em comento traz o princípio da explicabilidade como uma forma de garantir que os usuários e/ou agentes de inteligência artificial tenham conhecimento sobre quais dados são utilizados, a maneira como eles estão sendo processados, bem como o caminho percorrido pelo sistema para chegar a determinada conclusão ou previsão, a fim de que os usuários possam ter autonomia da decisão final sobre o tratamento de seus dados e também as

ferramentas necessárias para questionar decisões controversas, objetivando o equilíbrio entre avanço tecnológico e proteção dos direitos fundamentais.

Inicialmente, cumpre explicitar as definições pertinentes a este escopo trazidas pelo artigo 4º para melhor entendimento da normativa, a saber:

I – sistema de inteligência artificial: sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real;

II – fornecedor de sistema de inteligência artificial: pessoa natural ou jurídica, de natureza pública ou privada, que desenvolva um sistema de inteligência artificial, diretamente ou por encomenda, com vistas a sua colocação no mercado ou a sua aplicação em serviço por ela fornecido, sob seu próprio nome ou marca, a título oneroso ou gratuito;

III – operador de sistema de inteligência artificial: pessoa natural ou jurídica, de natureza pública ou privada, que empregue ou utilize, em seu nome ou benefício, sistema de inteligência artificial, salvo se o referido sistema for utilizado no âmbito de uma atividade pessoal de caráter não profissional;

IV – agentes de inteligência artificial: fornecedores e operadores de sistemas de inteligência artificial;

No que concerne à explicabilidade, o artigo 3º traz como princípios a serem atendidos no desenvolvimento, implementação e/ou uso de sistemas de inteligência artificial, a (ii) autodeterminação e liberdade de decisão e de escolha, (iii) participação humana no ciclo da inteligência artificial e supervisão humana efetiva, (iv) não discriminação, (vi) transparência, explicabilidade, inteligibilidade e auditabilidade, (ix) rastreabilidade das decisões, (x) prestação de cotas. Esses princípios decorrem direta ou indiretamente da explicabilidade e demonstram como determinada decisão obscura pode causar um verdadeiro efeito cascata em prejuízo sobre diversos conceitos principiológicos trazidos pela normativa, como auditabilidade, inteligibilidade, rastreabilidade, *etc.*

O artigo 5º traz os direitos a serem exercidos pelas pessoas afetadas por sistemas de inteligência artificial, elencando explicitamente o (ii) direito à

explicação sobre a decisão, recomendação ou previsão tomada por sistemas de inteligência artificial e também (iii) direito de contestar decisões ou previsões de sistemas de inteligência artificial que produzam efeitos jurídicos ou que impactam de maneira significativa os interesses do afetado, dispondo no parágrafo único que é de responsabilidade dos agentes de inteligência artificial informar de maneira clara e acessível os procedimentos necessários para exercício dos direitos elencados.

O direito à informação é ainda mais evidenciado através do artigo 7º, que explicita sobre quais aspectos devem ser prestadas informações claras previamente à contratação ou utilização do sistema, sendo eles:

Art. 7º Pessoas afetadas por sistemas de inteligência artificial têm o direito de receber, previamente à contratação ou utilização do sistema, informações claras e adequadas quanto aos seguintes aspectos:

- I – caráter automatizado da interação e da decisão em processos ou produtos que afetem a pessoa;
- II – descrição geral do sistema, tipos de decisões, recomendações ou previsões que se destina a fazer e consequências de sua utilização para a pessoa;
- III – identificação dos operadores do sistema de inteligência artificial e medidas de governança adotadas no desenvolvimento e emprego do sistema pela organização;
- IV – papel do sistema de inteligência artificial e dos humanos envolvidos no processo de tomada de decisão, previsão ou recomendação;
- V – categorias de dados pessoais utilizados no contexto do funcionamento do sistema de inteligência artificial;
- VI – medidas de segurança, de não-discriminação e de confiabilidade adotadas, incluindo acurácia, precisão e cobertura; e
- VII – outras informações definidas em regulamento.

A seguir, o artigo 8º consolida como norma o mais próximo do conceito de explicabilidade adotado, elencando que a pessoa afetada pode solicitar explicação sobre critérios, procedimentos e até fatores que levaram ao resultado, incluindo informações sobre (i) a racionalidade e a lógica do sistema, o significado e as consequências previstas de tal decisão para a pessoa afetada, (ii) o grau e o nível de contribuição do sistema de inteligência artificial para a tomada de decisões, (iii) os dados processados e a sua fonte, os critérios para a tomada de decisão e, quando apropriado, a sua ponderação, aplicados à situação da pessoa afetada, (iv) os mecanismos por meio dos quais a pessoa

pode contestar a decisão, (v) a possibilidade de solicitar intervenção humana, nos termos desta Lei. O parágrafo único amplia o acesso à informação determinando que as informações mencionadas devem ser fornecidas através de procedimento gratuito e facilitado, e ainda em linguagem compreensível em respeito aos níveis de alfabetização digital.

Desse modo, em consonância com a tendência mundial de regulamentação de um sistema de IA responsável e alinhado com os princípios éticos vigentes em sociedade, observa-se que o Projeto-lei em comento não se distanciou da necessidade de explicações cada vez mais acessíveis a serem prestadas à sociedade, demonstrando a preocupação do legislador no desenvolvimento de sistemas que possuam certa compreensibilidade não só por parte dos desenvolvedores e fornecedores como também dos usuários finais, a fim de que todo arcabouço ético seja satisfeito.

Portanto, a partir da análise do PL em comento infere-se que a regulamentação da IA no Brasil perpassa fortemente por normas éticas e de transparência, de modo que, o PL 2338/2023 reforça a importância da transparência e explicabilidade como pilares éticos no desenvolvimento e uso de sistemas de inteligência artificial no Brasil.

Inteligência Artificial Explicável (xAI)

Diante dos novos padrões exigidos para os sistemas de IA, tornou-se necessário o surgimento de uma abordagem que privilegiasse a explicabilidade do modelo, dando vida ao conceito de Inteligência Artificial Explicável, outrora denominada por xAI. Esse conceito refere-se a um conjunto de métodos que visam tornar a IA mais transparente e compreensível para os cidadãos, permitindo que os usuários possam entender as razões e a maneira como um sistema de AI encontrou determinada conclusão, decisão ou previsão, auxiliando ainda na eliminação de vieses e desenvolvimento de sistemas mais responsáveis na medida que seu funcionamento é mais controlado.

Nas palavras de Gunning *et al.* (2019, p.3), o objetivo de um sistema de IA explicável é tornar seu comportamento mais inteligível para os humanos por meio do fornecimento de explicações, de modo que, ainda consoante o autor, existem alguns princípios gerais que visam ajudar a criar esses sistemas de maneira mais eficaz e compreensível. Dentre eles, preceitua que o sistema

xAI deve ser capaz de explicar suas capacidades e entendimentos, explicar o que fez, o que está fazendo agora e o que acontecerá a seguir, e ainda divulgar as informações sobre as quais está agindo. Esse conceito traz uma boa noção do que vem a ser os sistemas de xAI em comento.

Traçando um paralelo com os sistemas de IA comuns, os sistemas de xAI têm seu foco completamente voltado para a transparência e interpretabilidade, priorizando a capacidade do modelo em explicar o processo da tomada de decisão, enquanto os primeiros mantêm seu foco integralmente no desempenho e precisão, sendo projetados para atingir os melhores resultados possíveis em suas tarefas, independente da explicação de como tais decisões são tomadas (SCHMELZER, 2019). Pode-se dizer que os sistemas de IA comuns funcionam como um tipo de caixa-preta, na qual os usuários geralmente não possuem informação sobre quais entradas específicas influenciam as saídas, enquanto o xAI busca adotar modelos explicáveis desde sua gênese ou utiliza de camadas de análise para tal objetivo, sendo razoável afirmar que a xAI atende melhor às regulamentações éticas e/ou de transparência mundialmente adotadas.

Antes de aprofundar o conteúdo abordado, faz-se necessária uma leve distinção entre os conceitos de transparência e explicabilidade abordados, ainda que por muitas vezes estejam entrelaçados de tal forma que tornam-se indissociáveis no contexto. Contudo, em abordagens mais rigidamente técnicas, a transparência relaciona-se com a clareza na arquitetura do modelo, o quanto seu funcionamento pode ser compreendido pelos usuários diretamente por meio de sua estrutura e funcionamento, enquanto a explicabilidade refere-se à capacidade do modelo em fornecer justificativas compreensíveis acerca de suas decisões e/ou previsões.

A explicabilidade diz respeito à capacidade de compreender e interpretar resultados apresentados por modelos de IA e resulta das ações e esforços feitos pela equipe que construiu o sistema em detalhar suas funções internas quando o modelo não pode ser devidamente interpretado diretamente. Assim, consoante a dificuldade apresentada para compreensão do modelo em cada caso particular, surge a necessidade de definição de ações que podem tornar o modelo mais compreensível. (MACOHIN, 2021)

Importante ressaltar que o PL 2.338/23 em comento, apesar de abordar princípios como da transparência e explicabilidade, deixa de definir o

significado de cada um desses termos perante a lei, não fazendo qualquer distinção entre os termos, limitando-se a explicitar que o projeto visa que os sistemas de IA sejam desenvolvidos de modo que seus resultados possam ser compreendidos pelo usuário. Isso reforça o quanto os conceitos citados estão diretamente entrelaçados e podem ser usados de maneira intercambiável em determinados contextos, de modo que a intenção legislativa primordial é a busca por uma IA ética e responsável.

Na prática, o nível de explicabilidade do modelo depende de fatores como a maneira de concepção do sistema, a escolha do algoritmo e a precisão do modelo, uma vez que modelos de acurácia alta podem ter sua compreensibilidade extremamente prejudicada devido a complexidade na elaboração de resultados, enquanto outros tipos de modelos, especialmente aqueles que tem o propósito de ser explicável desde o início do desenvolvimento do software, podem atingir uma melhor compreensibilidade pelos usuários.

Como exemplo de algoritmo complexo, podemos citar as redes neurais profundas (*deep learning*), que são um tipo de modelo projetado para simular o comportamento do cérebro humano, composto por várias camadas de neurônios interconectados que permitem que os sistemas aprendem padrões complexos em grandes volumes de dados. Esse tipo de algoritmo apresenta extrema dificuldade em explicar suas decisões em razão de fatores como a estrutura complexa em camadas, a natureza distribuída da representação, falta de causalidade explícita, etc. (SARKER, 2021)

Por outro lado, a árvore de decisão trata-se de um modelo conhecido pela sua fácil explicabilidade, principalmente em virtude de sua simplicidade e alta capacidade de representação virtual do processamento dos dados de entrada até a decisão e/ou previsão. Esse modelo utiliza-se de uma estrutura hierárquica que representa a lógica de regressão em “se-então” onde cada nó interno da árvore corresponde a uma condição baseada em uma característica, os ramos indicam os domínio de resultados da condição e as folhas representam os dados de saída como decisões finais ou previsões. Desse modo, esse modelo possui alta transparência nas decisões, permitindo que um usuário siga o caminho do nó raiz (condição de entrada) até uma folha (dado de saída), permitindo entender cada condição que levou a previsão concluída. (LOPES, 2023)

Além da escolha do modelo adequado a cada situação, também podem ser adotadas técnicas que favorecem a explicabilidade do modelo como LIME e SHAP, ambas podendo ser aplicadas independente do tipo de modelo utilizado. A primeira trabalha a partir da criação de um modelo linear aproximado para explicar as relações entre as variáveis de entrada e as de saída para cada caso concreto, sendo ideal para entender decisões individuais em modelos complexos. Em outras palavras, essa técnica gera explicações locais simplificando o modelo em torno de uma previsão específica, podendo fornecer compreensibilidade em modelos obscuros e/ou abstratos. (RIBEIRO, 2016)

O método SHAP visa calcular a contribuição de cada variável para uma determinada decisão, fornecendo explicações globais, sobre como as características impactam o modelo no conjunto de dados e também locais, explicando individualmente cada previsão e revelando de que maneira cada característica influenciou o resultado. Esse método é conhecido por ser interpretável e matematicamente robusto e tem ganho sobre os demais em virtude de sua interpretação ser bastante intuitiva, uma vez que os valores SHAP indicam a contribuição positiva ou negativa de cada característica para uma previsão com base em seu valor e sinal. (LUNDBERG, 2017)

Ainda assim, os desafios tecnológicos que permeiam o cumprimento integral da transparência e explicabilidade nos sistemas de inteligência artificial são diversos, dentre os quais, mas não somente (i) complexidade dos modelos, modelos avançados são altamente não lineares e tornam difícil mapear as entradas para saídas, (ii) custo computacional, especialmente em se tratando de modelos grandes ou conjuntos de dados de alta dimensionalidade, (iii) escalabilidade, sistemas distribuídos podem dificultar o cálculo eficiente de explicações do modelo, (iv) interpretação das explicações, mesmo as explicações podem ser difíceis de interpretar quando envolvem muitas variáveis ou cenários complexos, (v) viés e impacto ético, uma vez que técnicas de explicabilidade podem revelar vieses custosos para correção, como também as explicações tecnicamente corretas podem não ser facilmente compreendidas por usuários não técnicos e gerar interpretações equivocadas.

DISCUSSÃO

Dado o contexto trazido pela legislação em comento, observa-se que a Inteligência Artificial Explicável possui grandes ganhos em detrimento dos sistemas comuns, especialmente no que tange ao quesito da explicabilidade de seus resultados, característica imprescindível em se tratando do regular exercício dos direitos fundamentais relativos ao tratamento de dados da pessoa natural constitucionalmente garantidos e reafirmados pelo PL 2.338/23. Razões como mitigação de vieses, responsabilidade ética, facilidade de auditorias, redução de riscos e conformidade regulatória podem ser apontadas como determinantes para essa predileção.

Conforme já delineado anteriormente, o PL em comento não menciona qualquer distinção entre os conceitos de transparência e explicabilidade, de modo que, consoante interpretação, os dois conceitos podem ser definidos a partir do ponto de vista do usuário, sobre o quanto determinado indivíduo consegue compreender sobre como o sistema chegou a determinado resultado. Isso pode ser visto como uma vantagem positiva, dado que, atendida a intenção do legislador de fornecer parâmetros de explicação aos usuários de inteligência artificial, a responsabilidade ética do sistema está configurada para este requisito.

Contudo, ainda que o sistema seja devidamente classificado como explicável, a implantação de um sistema de IA explicável também esbarra em algumas barreiras sociais que podem prejudicar seu funcionamento, dentre os quais a diferença nos níveis de alfabetização digital, divergências culturais, conflito de confiança, questões éticas e vieses sociais, as quais devem ser analisados caso a caso para adoção das ferramentas para superar esses imbróglios, vez que a explicação fornecida pelo sistema pode ser particularizada caso a caso.

Para além, apesar da explicabilidade ser uma propriedade proveitosa, nem sempre pode ser classificada como uma necessidade, existindo alguns pontos que devem ser analisados, vez que demandar que o sistema de IA explique cada decisão particular pode causar uma menor eficiência do sistema, podendo gerar resultados que apesar de explicáveis, não são o resultado mais adequado para a entrada fornecida (NORVIG, 2019).

No que tange ao setor público, os esforços para tornar a explicabilidade da IA mais acessível aos usuários merece atenção, uma vez que as decisões

tomadas impactam diretamente nas políticas públicas e confiança entre governo e cidadãos. Por essa razão, o PL 2.338/23 é severo ao delinear as ferramentas para que a IA seja cada vez mais acessível ao cidadão e este possa ter controle das decisões que são tomadas com seus dados. Também é necessária a atenção sobre de que forma o tratamento de dados é feito, prestigiando seu direito de regresso em caso de decisão manifestamente enviesada, de maneira que a compreensibilidade do algoritmo torna o exercício desses direitos extremamente facilitado.

Nesse sentido, acredita-se que a necessidade de explicabilidade de modelos utilizados pela máquina pública prepondera sobre uma maior eficiência do algoritmo, uma vez que para além do PL em comento, a transparência é um pilar fundamental do estado democrático de direito. Faz-se necessário a ponderação entre os modelos disponíveis para atender cada caso em particular, uma vez que sistemas não considerados de alto risco podem utilizar-se de modelos mais complexos em privilégio a eficiência dos resultados buscados.

Ainda pode-se utilizar da chamada hibridização de modelos, que nada mais é do que a utilização de abordagem híbrida que combina modelos simples e explicáveis com outros modelos mais complexos e eficientes, de modo que o modelo simples pode clarificar as decisões tomadas enquanto o modelo complexo mantém a alta performance para o resultado almejado. Como exemplo, podemos prever a elegibilidade de determinado benefício social, onde o modelo complexo faria a categorização inicial enquanto o modelo explicável é utilizado para auditar e justificar as decisões tomadas.

CONCLUSÃO

Conforme exaustivamente discutido no presente manuscrito, restou clarificado que o Projeto de Lei nº 2338/2023 consolida a explicabilidade como elemento essencial para a ética e o desenvolvimento de uma inteligência artificial responsável. Ao estabelecer a transparência e a inteligibilidade como princípios fundamentais, o PL assegura que os sistemas de IA sejam compreensíveis e auditáveis, promovendo a confiança pública e garantindo que as tecnologias operem de forma alinhada aos valores éticos e aos direitos fundamentais. Essa abordagem reconhece que a explicação não é apenas um requisito técnico,

mas também uma exigência ética imprescindível para a adoção segura e justa da IA em diversos contextos.

Para alcançar esses objetivos, torna-se necessária uma análise caso a caso que equilibre a eficiência e a explicabilidade dos sistemas de IA, priorizando a maior compreensão possível dentro do modelo adotado a fim de assegurar os direitos do usuário. Essa abordagem permite ajustar a transparência às especificidades de cada aplicação, sem comprometer o desempenho ou a confiabilidade do sistema. Assim, ao buscar maximizar a compreensão e a rastreabilidade, o desenvolvimento responsável de IA pode atender tanto às demandas éticas quanto aos desafios tecnológicos e sociais, pavimentando o caminho para uma adoção mais sustentável e confiável dessas tecnologias.

REFERÊNCIAS

ARRIETA, Alejandro Barredo *et al.* **Explainable Artificial Intelligence (XAI)**: Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, v. 58, 2020. p. 82-115.

BRASIL. PL 21/2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 19 de dezembro de 2024.

BRASIL. PL 2338/2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 21 de dezembro de 2024.

GOVERNO FEDERAL. Plano Brasileiro de IA terá supercomputador e investimento de R\$ 23 bilhões em quatro anos. Ministério da Ciência, Tecnologia e Inovações, 2024. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/07/plano-brasileiro-de-ia-tera-supercomputador-e-investimento-de-r-23-bilhoes-em-quatro-anos?utm_source=chatgpt.com. Acesso em: 21 de dezembro de 2024.

GUNNING, David *et al.* XAI— Explainable artificial intelligence. **Science Robotics**, v. 4, n. 37, 2019.

IA RESPONSÁVEL. Normas técnicas que tratam de Inteligência Artificial. IA Responsável, 07 jun. 2023. Disponível em: <https://www.iaresponsavel.com.br/2023/06/07/normas-tecnicas-que-tratam-de-inteligencia-artificial/>. Acesso em: 20 de dezembro de 2024.

LOPES, André. Árvores de Decisão: Algoritmos Baseados em Árvores. **BRAINS**, 2023. Disponível em: <https://brains.dev/2023/arvores-de-decisao-algoritmos-baseados-em-arvores/>. Acesso em: 23 de dezembro de 2024.

LUNDBERG, Scott M.; LEE, Su-In. **A Unified Approach to Interpreting Model Predictions**. arXiv preprint, 2017. Disponível em: <https://arxiv.org/abs/1705.07874>. Acesso em: 26 de dezembro de 2024.

MCCARTHY, John *et al.* **A proposal for the Dartmouth Summer Research Project on Artificial Intelligence**, August 31, 1955. *AI Magazine*, v. 27, n. 4, p. 12-12, 2006.

NEW WORLD AI. **Google's approach to Artificial Intelligence Machine Learning**: Peter Norvig. New World AI, 2019. Disponível em: <https://www.newworldai.com/googles-approach-to-artificial-intelligence-machine-learning-peter-norvig/>. Acesso em: 10 de dezembro de 2024.

NORVIG, Peter; RUSSELL, Stuart. **Inteligência artificial**. Rio de Janeiro: Grupo GEN, 2013.

OECD. OECD Framework for the Classification of AI Systems. **OECD Digital Economy Papers**, Nº. 323, OECD Publishing, Paris, 2022.

OECD. Recommendation of the Council on Artificial Intelligence. 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 11 de dezembro de 2024.

BENEDITO, Luana Maria. Brasil propõe plano de investimento de US\$ 4 bilhões em IA. **Reuters**, 30 jul. 2024. Disponível em: <https://www.reuters.com/technology/artificial-intelligence/brazil-proposes-4-billion-ai-investment-plan-2024-07-30/>. Acesso em: 11 de dezembro de 2024.

RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTRIN, Carlos. **"Why Should I Trust You?"**: Explaining the Predictions of Any Classifier. arXiv preprint, 2016. Disponível em: <https://arxiv.org/abs/1602.04938>. Acesso em: 27 de dezembro de 2024.

SARKER, Iqbal H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. **SN Computer Science**, v. 2, n. 6, p. 1-20, 2021. Disponível em: <https://link.springer.com/article/10.1007/s42979-021-00815-1>. Acesso em: 20 de dezembro de 2024.

SCHMELZER, Ron. Understanding Explainable AI. **Forbes**, 23 jul. 2019. Disponível em: <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/?sh=1e04b2bd7c9e>. Acesso em: 21 de dezembro de 2024.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de março de 2024 que estabelece regras harmonizadas sobre inteligência artificial (Regulamento de Inteligência Artificial) e altera determinados atos legislativos da União. **Jornal Oficial da União Europeia**, L 295, 14 de março de 2024, p. 1-60. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32024R1689>. Acesso em: 17 de dezembro de 2024.

UNESCO. Recomendação sobre a ética da inteligência artificial. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000380455_por. Acesso em: 18 de dezembro de 2024.

WACHOWICZ, Marcos. **Regulação da Inteligência Artificial**: Garantindo um Futuro Ético e Inclusivo no Brasil. GEDAI/UFPR, 2023. Disponível em: <https://gedai.ufpr.br/regulacao-da-inteligencia-artificial-garantindo-um-futuro-etico-e-inclusivo-no-brasil/>. Acesso em: 23 de dezembro de 2024.

04

IA Y ROBOTICS APLICADOS AL ANÁLISIS FINANCIERO EN UNA EMPRESA DEL SECTOR PESQUERO: UNA APROXIMACIÓN DESDE LA CONTABILIDAD GERENCIAL

Patricia Cáceres Ortega

Sonia Jackeline Miranda Avalos

Juan Alberto Villagómez Chinchay

Edwin Sixto Moncada Ochoa

Víctor Dante Ataupillco Vera

RESUMEN

El presente estudio tiene como objetivo establecer un conjunto de análisis sobre los estados financieros y el rubro de existencias bajo los alcances de la contabilidad gerencial en una empresa del sector pesquero peruano. Para ello, se propone una metodología aplicativa basada en el análisis de información financiera y contable documental, utilizando técnicas avanzadas de análisis financiero y automatización robótica, así como de la inteligencia artificial. La investigación se centra en una empresa con más de 40 años de trayectoria en la fabricación y comercialización de productos para la industria pesquera y otros sectores, donde se han implementado herramientas de análisis financiero y control de inventarios. Los resultados evidencian que el uso de tecnologías como el software IDEA permite optimizar la evaluación de riesgos financieros y operacionales mediante la identificación de patrones en los estados financieros y en el movimiento de inventarios. Se detectaron inconsistencias en el registro de comprobantes de almacén, operaciones realizadas en días no laborables y variaciones significativas en costos unitarios, lo que representa áreas de mejora en la gestión financiera y operativa. En conclusión, la aplicación de herramientas de contabilidad gerencial apoyadas en inteligencia artificial y tecnología robótica con análisis automatizado mejora significativamente la eficiencia del control financiero y reduce los riesgos asociados a la administración de inventarios en el sector pesquero. Se recomienda continuar explorando la integración de modelos predictivos y herramientas de inteligencia artificial para optimizar la toma de decisiones empresariales.

Palabras clave: contabilidad gerencial; análisis financiero; sector pesquero; automatización robótica; gestión de riesgos.

INTRODUCCIÓN

En el ámbito empresarial, la contabilidad gerencial juega un papel fundamental en la toma de decisiones estratégicas al proporcionar herramientas de análisis financiero que permiten evaluar la eficiencia operativa y la rentabilidad de las organizaciones. En el sector pesquero peruano, donde las fluctuaciones del mercado, la variabilidad en las capturas y la gestión de recursos representan desafíos constantes, el adecuado manejo de los estados financieros y del rubro de existencias resulta crucial para garantizar la sostenibilidad y competitividad de las empresas.

En la actualidad, el uso de herramientas informáticas especializadas en contabilidad y gestión financiera se ha convertido en un recurso indispensable para el análisis de los estados financieros y la gestión del rubro de existencias en las empresas. Software como SAP ERP, QuickBooks, Contasol, Microsoft Power BI y Excel avanzado permiten procesar grandes volúmenes de datos contables, automatizar reportes financieros y generar indicadores clave para la toma de decisiones. En el caso específico del rubro de existencias, herramientas de gestión de inventarios como Odoo, NetSuite y Softland facilitan el control en tiempo real de los insumos y productos terminados, optimizando los niveles de stock y reduciendo pérdidas por obsolescencia o deterioro. Además, mediante la integración de algoritmos de análisis predictivo y dashboards interactivos, estas plataformas permiten visualizar tendencias, evaluar la rentabilidad de las operaciones y proyectar escenarios financieros futuros con mayor precisión. La implementación de estas tecnologías en el sector pesquero y otras industrias mejora la eficiencia operativa, minimiza riesgos financieros y contribuye a una gestión más estratégica de los recursos empresariales.

Este trabajo tiene como objetivo establecer un conjunto de análisis sobre los estados financieros y el rubro de existencias bajo los alcances de la contabilidad gerencial en una empresa del sector pesquero peruano. La propuesta aplicativa es la utilización de algoritmos robóticos que realicen el seguimiento de los rubros identificados para el análisis. A través de esta evaluación, se busca proporcionar información relevante que facilite la optimización de costos, el control de inventarios y la mejora en la eficiencia operativa, elementos esenciales para la rentabilidad y estabilidad financiera de la empresa analizada.

REVISIÓN DE LITERATURA

Contabilidad Gerencial (CG)

La Contabilidad Gerencial es una función crítica dentro de las organizaciones, que proporciona información esencial para la toma de decisiones, la planificación y el control. Implica la recopilación, análisis y presentación de datos financieros para ayudar a los directivos a tomar decisiones informadas.

Funciones clave

- Apoyo a la toma de decisiones. La contabilidad gerencial proporciona información financiera detallada que ayuda a los directivos a tomar decisiones tácticas y estratégicas. Esto incluye actividades operativas, financieras y de información (CALOTĂ, 2021; DIAKOV, 2010).
- Gestión de costes. Desempeña un papel crucial en la gestión de los costes de producción y la optimización del sistema de producción para garantizar la eficiencia y la competitividad (ALSHARARI, 2024; CALOTĂ, 2021).
- Gestión estratégica. Con el tiempo, la contabilidad gerencial ha evolucionado para apoyar la gestión estratégica, vinculando la información financiera con los objetivos estratégicos y contribuyendo así a la aplicación eficaz de las estrategias (ADAMOVA; KHABIB; TEPLYAKOVA, 2020; PICKERING, 2022).
- Medición del rendimiento. Se utiliza para medir y controlar el rendimiento organizativo, garantizando que la empresa cumple sus objetivos financieros y operativos (HEIDARI, 2012; JACKOVÁ, 2015).

Evolución y prácticas modernas

- Evolución histórica. La contabilidad gerencial ha evolucionado significativamente desde sus inicios en la industria textil de Nueva Inglaterra y en la armería de Springfield hasta los avanzados sistemas utilizados en la actualidad. Esta evolución ha sido impulsada por la necesidad de disponer de información financiera más precisa y relevante (ALSHARARI, 2024; MICULESCU, 2013).

- Contabilidad de gestión estratégica. La contabilidad de gestión moderna ha integrado las prácticas de gestión estratégica, centrándose en los objetivos a largo plazo y en la dirección estratégica global de la empresa (PICKERING, 2022; VLADIMIROVNA GALAUTDINOVA, 2019).
- Big Data y tecnología: La incorporación de big data y analítica avanzada ha mejorado aún más las capacidades de la contabilidad de gestión, permitiendo una previsión más precisa y una mejor toma de decisiones (ABDULLAH; SANUSI; SAVITRI, 2022; CHÁVEZ-DÍAZ, 2024; CHIPRIYANOVA; KRASTEVA-HRISTOVA, 2023).

Herramientas y técnicas

- Sistemas de cálculo de costes. Para asignar los costes con precisión y gestionar los gastos se utilizan diversos sistemas de cálculo de costes, como el cálculo de costes por puesto de trabajo, el cálculo de costes por proceso y el cálculo de costes por actividades (ALSHARARI, 2024).
- Presupuestación y previsión: Son herramientas esenciales para planificar y controlar los recursos financieros, garantizando que la organización se mantiene en la senda de cumplir sus objetivos financieros (ALESANI, 2023; CALOTĂ, 2021).
- Seguimiento del rendimiento. Técnicas como el Análisis de Rentabilidad de Clientes (CPA) ayudan a comprender la rentabilidad de los distintos segmentos de clientes y a tomar decisiones estratégicas con conocimiento de causa (LUEG; ILIEVA, 2024).

Retos y consideraciones

- Integración con objetivos estratégicos. Uno de los principales retos es alinear las prácticas de contabilidad de gestión con los objetivos estratégicos de la organización. Para ello, es necesario comprender en profundidad tanto los objetivos financieros como los estratégicos (ADAMOVA; KHABIB; TEPLYAKOVA, 2020; LUEG; ILIEVA, 2024).

- Adaptación a los cambios del mercado. La contabilidad de gestión debe ser flexible y adaptarse al entorno dinámico del mercado, proporcionando información pertinente y oportuna para la toma de decisiones (ADAMOVA; KHABIB; TEPLYAKOVA, 2020; HENG; FIEDLER; KANDUNIAS, 2005).
- Sostenibilidad y RSE. La contabilidad de gestión moderna también tiene en cuenta la sostenibilidad de los aspectos sociales, económicos y medioambientales, garantizando que las empresas actúen de forma responsable (ONCIOIU, 2021).

Conviene precisar que la Contabilidad Gerencial es una subcategoría dentro de la Contabilidad de Gestión. La primera se enfoca en análisis contables internos, la segunda integra herramientas de gestión, control y estrategia empresarial.

En la práctica, ambos términos suelen superponerse, pero la contabilidad de gestión (Management Accounting) es un concepto más amplio que incluye a la contabilidad gerencial.

Por lo obtenido de la revisión de literatura, la contabilidad de gestión es una parte indispensable de la gestión empresarial moderna, ya que proporciona la información necesaria para la toma de decisiones, la planificación y el control eficaces. Su evolución e integración con las prácticas de gestión estratégica la han convertido en una herramienta vital para alcanzar los objetivos de la organización y mantener la competitividad en un entorno de mercado dinámico.

Sector Pesquero Peruano (SPP)

La competitividad del sector pesquero depende de su capacidad para producir y comercializar productos alineados con la demanda del mercado, considerando las preferencias del consumidor, manteniendo altos estándares de calidad y asegurando precios competitivos. Para fortalecer esta competitividad, es esencial un análisis detallado de las cadenas productivas, dado que su solidez depende del eslabón más vulnerable.

El sector pesquero peruano es una de las principales fuentes de ingresos y empleo en el país, contribuyendo significativamente al Producto Bruto Interno (PBI) y a las exportaciones (ARANDA, 2009; ARIAS SCHREIBER;

ÑIQUEN; BOUCHON, 2011; GOZZER-WUEST; ALONSO-POBLACIÓN; TINGLEY, 2021). Su importancia radica en la diversidad de especies marinas que ofrece el litoral peruano, destacando la anchoveta como la principal especie para la producción de harina y aceite de pescado (ARANDA, 2009; TICSE-VILLANUEVA *et al.*, 2021). Sin embargo, la industria enfrenta desafíos relacionados con la sobreexplotación de recursos, regulaciones ambientales y la variabilidad climática causada por fenómenos como El Niño, que afectan las capturas y, por ende, la estabilidad del sector (ARIAS SCHREIBER; ÑIQUEN; BOUCHON, 2011; COAYLA *et al.*, 2023).

Desde una perspectiva económica, el sector pesquero peruano se divide en dos grandes segmentos: la pesca industrial y la pesca artesanal. La pesca industrial se enfoca en la extracción masiva de especies destinadas a la exportación, principalmente para la producción de harina y aceite de pescado. En cambio, la pesca artesanal, que es vital para la seguridad alimentaria y la economía de las comunidades costeras, enfrenta problemas de infraestructura, acceso a financiamiento y sostenibilidad. La informalidad en este subsector sigue siendo un obstáculo para su desarrollo, afectando la trazabilidad y el control de las capturas (JAMES *et al.*, 2023; SUÁREZ-CASTAÑEDA, 2025).

En términos de sostenibilidad y gestión ambiental, el país ha implementado regulaciones estrictas para garantizar el uso responsable de los recursos marinos. La cuota de pesca para la anchoveta, el control de las temporadas de captura y las medidas de fiscalización buscan evitar la sobrepesca y la degradación del ecosistema marino. No obstante, persisten desafíos relacionados con la pesca ilegal y el cumplimiento de las normativas, lo que requiere un esfuerzo conjunto entre el gobierno, las empresas y las comunidades pesqueras para fortalecer la gobernanza del sector (GOZZER-WUEST; ALONSO-POBLACIÓN; TINGLEY, 2021; TICSE-VILLANUEVA *et al.*, 2021).

El Tratado de Libre Comercio Perú-China ha facilitado exportaciones significativas, pero las inversiones extranjeras no han beneficiado sustancialmente a la fuerza laboral local ni han conducido a avances económicos y tecnológicos significativos (MALDONADO-CUEVA; CRIADO-DÁVILA; FERNANDEZ-BEDOYA, 2024).

La capacidad de las empresas pesqueras para competir en el mercado está determinada por el nivel de sofisticación de sus operaciones y estrategias,

así como por la calidad del entorno microeconómico en el que operan. La competitividad no solo implica eficiencia interna, sino también la optimización y el fortalecimiento de toda la cadena de valor.

De acuerdo con Porter (1985), el análisis de la ventaja competitiva de una actividad requiere una evaluación de su cadena de valor, ya que permite identificar las actividades clave que la caracterizan. En este contexto, la cadena de valor se define como el conjunto de actividades necesarias para la producción de bienes o servicios en el sector pesquero, dentro de un entorno con características específicas.

En el sector pesquero, la cadena de valor se compone de cuatro eslabones fundamentales. La primera etapa corresponde a la captura o extracción de recursos hidrobiológicos, realizada mediante la pesca tradicional o la acuicultura. La segunda fase, el procesamiento, consiste en la transformación de la materia prima en productos finales de diverso tipo. El tercer eslabón, la distribución, abarca las actividades vinculadas a la comercialización y logística de los productos pesqueros. Finalmente, el consumo representa la etapa en la que los productos llegan a los consumidores finales.

Cada una de estas fases involucra a distintos agentes económicos que desempeñan funciones específicas para generar valor agregado a lo largo de la cadena productiva. La optimización de estos procesos no solo contribuye a mejorar la competitividad del sector, sino que también promueve un consumo responsable y sostenible de los recursos pesqueros.

Figura 1 - Esquema de producción y consumo en sector pesca.



Fuente: Tomado de Galarza & Kámiche (2015)

El sector pesquero peruano enfrenta diversas limitaciones en la competitividad de su cadena de valor, lo que representa un desafío, pero también una oportunidad para su desarrollo. Entre las principales áreas de mejora se encuentran el manejo eficiente de los estados financieros y la gestión de inventarios en las empresas del sector. Asimismo, resulta fundamental impulsar la Investigación, el Desarrollo y la Innovación (I+D+i) como un factor clave para incrementar la competitividad.

En este sentido, el uso de herramientas informáticas especializadas en contabilidad y gestión financiera representa una estrategia eficaz para mejorar la eficiencia operativa y promover un crecimiento empresarial más equitativo y sostenible.

El análisis integral de la cadena de valor pesquero y de los estados financieros permite identificar estrategias para optimizar los procesos, mejorar la eficiencia operativa y aumentar la rentabilidad de las empresas del sector. Además, este enfoque fortalece la capacidad competitiva de las empresas pesqueras, facilitando su adaptación a los desafíos del mercado y la identificación de nuevas oportunidades de desarrollo y crecimiento.

Por lo indicado, el avance tecnológico y la digitalización están jugando un papel clave en la modernización del sector pesquero peruano. La incorporación de herramientas de análisis financiero, gestión de inventarios y automatización de procesos ha mejorado la eficiencia operativa de las empresas pesqueras, permitiéndoles optimizar costos y mitigar riesgos financieros. A pesar de estos avances, el sector debe seguir evolucionando en aspectos de innovación, sostenibilidad y diversificación de mercados para garantizar su competitividad en el contexto global.

METODOLOGÍA

Estudio de tipo documental y aplicativo sobre información estratégica, financiera y contable de una empresa del sector pesquero. Se utilizan técnicas de análisis documental sobre las memorias de la empresa, así como libros contables como el Libro Diario y el Registro de Inventario Permanente Valorizado.

Se consideró una empresa familiar con más de 40 años de experiencia dedicada a la fabricación y comercialización de redes, cordeles y cabos destinados principalmente para el sector pesca. Sin embargo, sus clientes abarcan diversos sectores como deporte, agropecuario y avícola. Mediante el uso progresivo de

tecnología avanzada, se caracteriza por su compromiso con la productividad y la competitividad.

La investigación utilizará el complemento robótico del software IDEA de Caseware International, para lo cual se realizará un mapeo de la información disponible, se establecerán los riesgos a evaluar y se generarán los reportes apropiados de análisis y seguimiento.

RESULTADOS

Análisis de Estados Financieros

Se realizaron evaluaciones sobre el Libro Diario de la empresa y los estados financieros del período 2024, con el uso de la inteligencia artificial. Se presenta una relación de ratios calculados, ver tabla 1.

Tabla 1 - Ratios financieros.

Ratio	Fórmula	Valores S/	Valor Calculado
Razón Corriente	Activo Corriente / Pasivo Corriente	15,347,934.11 / 13,697,714.55	1.120474081
Prueba Ácida	(Activo Corriente - Inventarios) / Pasivo Corriente	(15,347,934.11 - 4,814,346.18) / 13,697,714.55	0.769003318
ROA (Return on Assets)	Utilidad Neta / Activo Total	1,902,351.54 / 35,564,453.91	0.05349025
ROE (Return on Equity)	Utilidad Neta / Patrimonio Neto	1,902,351.54 / 14,220,270.98	0.133777446
Margen de Utilidad	Utilidad Neta / Ventas	1,902,351.54 / 23,521,649.31	0.080876622
Rotación de Activos	Ventas / Activo Total	23,521,649.31 / 35,564,453.91	0.661380865
Rotación de Cuentas por Cobrar	Ventas / Cuentas por Cobrar	23,521,649.31 / 4,663,022.36	5.044292629
Rotación de Inventarios	Costo de Ventas / Inventario Total	14,745,669.79 / 4,814,346.18	3.062860301
Razón de Endeudamiento	Pasivo Total / Activo Total	21,344,182.93 / 35,564,453.91	0.600154946
Razón Deuda/ Patrimonio	Pasivo Total / Patrimonio Neto	21,344,182.93 / 14,220,270.98	1.500968791

Fuente: En Anexo 1, se muestra el procedimiento de generación mediante IA.

Se comprobó los resultados obtenidos siendo eficiente el proceso de generación de ratios. El análisis de la empresa a través de los ratios financieros

permite evaluar su liquidez, rentabilidad, eficiencia operativa y nivel de endeudamiento. A continuación, se interpreta la salud financiera de la empresa con base en los resultados obtenidos.

1. Liquidez

- Razón Corriente (1.12): La empresa tiene 1.12 unidades monetarias en activos corrientes por cada unidad monetaria en pasivos corrientes, lo que indica que puede cubrir sus obligaciones a corto plazo. Sin embargo, este valor no es demasiado holgado, lo que sugiere que la empresa debe gestionar con cuidado su capital de trabajo.
- Prueba Ácida (0.77): Al descontar los inventarios, la capacidad de pago inmediato disminuye significativamente. Con un valor de 0.77, la empresa podría enfrentar dificultades para cubrir sus pasivos corrientes sin recurrir a la venta de inventarios. Esto indica que una parte importante de su liquidez está atada a existencias, lo que podría ser riesgoso en caso de una caída en la demanda.

2. Rentabilidad

- ROA (5.35%): La rentabilidad sobre los activos es relativamente baja, lo que indica que la empresa está generando un rendimiento moderado en relación con los recursos que posee.
- ROE (13.38%): La rentabilidad sobre el patrimonio es superior al ROA, lo que muestra que la empresa está logrando un retorno aceptable sobre la inversión de los accionistas.
- Margen de Utilidad (8.09%): La empresa obtiene 8.09 centavos de utilidad por cada sol en ventas, lo que indica un margen razonable. Sin embargo, un análisis más profundo de los costos podría ayudar a identificar oportunidades de mejora.

3. Eficiencia Operativa

- Rotación de Activos: Indica qué tan bien la empresa utiliza sus activos para generar ingresos. Un valor bajo podría sugerir que los activos no están siendo utilizados de manera óptima.

- Rotación de Cuentas por Cobrar: Refleja la eficiencia en la gestión de créditos a clientes. Si la rotación es baja, la empresa podría estar enfrentando problemas de cobranza.
- Rotación de Inventarios: La velocidad con la que se venden y reponen los inventarios impacta la liquidez. Una baja rotación podría indicar problemas de obsolescencia o sobreacumulación de stock.

4. Endeudamiento

- Razón de Endeudamiento (60.02%): Más del 60% de los activos están financiados con deuda, lo que indica un nivel de apalancamiento moderado. Aunque no es alarmante, la empresa debe mantener un control sobre su carga financiera para evitar riesgos en caso de un deterioro económico.
- Razón Deuda/Patrimonio (1.50): Existe 1.50 unidades monetarias de deuda por cada unidad de patrimonio, lo que sugiere que la empresa depende significativamente del financiamiento externo. Aunque este nivel de endeudamiento puede ser manejable, es crucial que la empresa tenga flujos de efectivo sólidos para cumplir con sus obligaciones financieras.

Desde una perspectiva de contabilidad gerencial, la empresa tiene una posición financiera aceptable, pero con áreas de mejora. La liquidez es justa y depende en gran parte de los inventarios, lo que podría representar un riesgo en escenarios de menor demanda. La rentabilidad es positiva, pero podría optimizarse a través de una mejor gestión de costos y una mayor eficiencia en el uso de activos. Finalmente, el nivel de endeudamiento es moderado, lo que indica que la empresa debe monitorear su estructura de capital para evitar problemas financieros a largo plazo.

Para fortalecer su posición financiera, la empresa podría:

- Optimizar su gestión de inventarios para mejorar la liquidez.
- Aumentar la rotación de cuentas por cobrar para mejorar el flujo de caja.
- Evaluar estrategias para reducir costos y mejorar la rentabilidad.
- Revisar su estructura de financiamiento para minimizar el riesgo de sobreendeudamiento.

En general, la empresa está en una situación financiera estable, pero debe enfocarse en la gestión eficiente de su capital de trabajo y en mejorar su rentabilidad para asegurar un crecimiento sostenible.

Análisis de inventarios

Se han identificado los siguientes temas de análisis continuo de información sobre el Registro de Inventarios Permanente Valorizado, se incluye la codificación correspondiente, ver tabla 2.

Tabla 2 - Temas de análisis en inventarios

Nro.	Tema	Descripción	Código
01	Días no laborales	Despacho de inventarios en días no laborales (fin de semana, feriados)	Reporte_01
02	Costo de ventas	Seguimiento detallado del Costo de Ventas	Reporte_02
03	Estacionalidad	Análisis histórico de salidas del inventario	Reporte_03
04	Duplicados de comprobantes	Registros sospechosos de duplicados en almacén	Reporte_04
05	Prueba de fraude	Comportamiento anómalo de valores absolutos de registro de inventarios	Reporte_05
06	Saldo Inicial	Análisis de saldos iniciales	Reporte_06
07	Movimiento de items	Análisis de items con mayor movimiento y rotación	Reporte_07
08	Movimiento de tipo de docs	Análisis de ingresos y salidas por documento sustentatorio	Reporte_08
09	Costos unitarios	Evaluación y control de precios	Reporte_09
10	Precios y costos unitarios	Precios de ventas por debajo del costo unitario	Reporte_10

Producto de la ejecución del algoritmo se generaron las siguientes tablas, ver tabla 3.

Tabla 3 - Resultados de proceso robótico sobre Inventarios.

Nro.	Tipo	Nombre	Descripción
01	Tabla	Reporte_01_fds	Registros que cumplen el criterio de operaciones en fin de semana
02	Tabla	Reporte_02_detalle_CVentas	Registros de operaciones de venta en kardex
03	Tabla	Reporte_03_Estacionalidad	Tabla salidas de almacén por items mensual
04	Tabla	Reporte_04_Duplicados	Registros sospechosos de duplicados en almacén
05	Tabla	Reporte_05_Benford_primer_digito	Prueba Benford de un dígito
06	Tabla	Reporte_06_Saldo_Inicial	Resumen por item de saldos iniciales por item

Nro.	Tipo	Nombre	Descripción
07	Tabla	Reporte_07_Codigo	Resumen de movimientos por item
08	Tabla	Reporte_08_Tdoc	Análisis de ingresos y salidas por documento sustentatorio
09	Tabla	Reporte_09_ranking_item	Ranking de ítems vendidos
10	Tabla	Reporte_09_item1	Resumen de ítem primer lugar
11	Tabla	Reporte_09_item2	Resumen de ítem segundo lugar
12	Tabla	Reporte_09_item3	Resumen de ítem tercer lugar
13	Tabla	Reporte_09_detalle_item1	Detalle de ítem primer lugar
14	Tabla	Reporte_09_detalle_item2	Detalle de ítem segundo lugar
15	Tabla	Reporte_09_detalle_item3	Detalle de ítem tercer lugar
16	Tabla	Reporte_09_Cunit_item1	Estadísticas de costos unitarios de ítem primer lugar
17	Tabla	Reporte_09_Cunit_item2	Estadísticas de costos unitarios de ítem segundo lugar
18	Tabla	Reporte_09_Cunit_item3	Estadísticas de costos unitarios de ítem tercer lugar
19	Tabla	Reporte_09_Cunit	Estadísticas de costos unitarios de los primeros tres ítems

En la figura 2 se presenta una vista de los resultados en el Software IDEA, en Anexo 2 se presentan los algoritmos de programación.

Figura 2 - Vista de tablas resultantes en Kardex.

Explorador de archivos

Nombre	Registros	Tamaño	Modificado
KG-Hoja1	253	156	20/02/2025 13:39:07
Reporte_01_fds	2	32	20/02/2025 13:39:02
Reporte_02_detalle_CVentas	138	76	20/02/2025 13:39:02
Reporte_03_Estacionalidad	3	10	20/02/2025 13:39:04
Reporte_04_Duplicados	141	66	20/02/2025 13:39:05
Reporte_05_Benford_primer_digito	9	28	20/02/2025 13:39:01
Reporte_06_Saldo_Inicial	3	17	20/02/2025 13:38:59
Reporte_07_Codigo	3	26	20/02/2025 13:39:05
Reporte_09_ranking_item	3	31	20/02/2025 13:39:07
Reporte_09_item1	1	30	20/02/2025 13:39:05
Reporte_09_item2	1	30	20/02/2025 13:39:06
Reporte_09_item3	1	30	20/02/2025 13:39:07
Reporte_08_Tdoc	11	25	20/02/2025 13:39:00
Reporte_09_detalle_item1	101	90	20/02/2025 13:39:06
Reporte_09_Cunit_item1	1	43	20/02/2025 13:39:08
Reporte_09_Cunit	3	37	20/02/2025 13:39:08
Reporte_09_detalle_item2	53	74	20/02/2025 13:39:07
Reporte_09_Cunit_item2	1	38	20/02/2025 13:39:08
Reporte_09_detalle_item3	99	85	20/02/2025 13:39:08
Reporte_09_Cunit_item3	1	38	20/02/2025 13:39:08

Entre los hallazgos obtenidos claves para la toma de decisiones se presentan los siguientes.

a. Control de fin de semana

Figura 3 - Análisis de fin de semana.

Reporte_01_fds							
	NCODIGO	ITEM	COMPRBTE	FECHA	ALMACEN	CANT_SALI	VAL_TOTAL
1	Red nylon 210/18 x 1 1/2" x 200 M.A., negro - (KG)	25	201-012880	31/03/2024	ALMA01	227.00	-4,619.45000
2	Red nylon 210/18 x 1 1/2" x 200 M.A., negro - (KG)	46	201-013295	30/06/2024	ALMA01	290.00	-5,910.20000

En la figura 3, se presenta que en el período 2024 ha habido atenciones en días no laborables por la empresa. El monto valorizado supera los S/10,000.

En efecto se realizaron en el sistema dos salidas de Red nylon 210/18 x 1 1/2" x 200 M.A., color negro en días no laborables. La primera fue de 227 Kg el día 31 de marzo y la segunda de 290 Kg el día 30 de junio. En ambos casos las salidas quedaron registradas en el sistema un día domingo.

En el primer caso, un cliente realizó un depósito el domingo 31 marzo, correspondiente a un pedido de la red mencionada. El día lunes 1º de abril, al realizarse el cierre del mes de marzo se procedió a la facturación de la red solicitada, lo que generó automáticamente la salida en el sistema de almacén, a pesar de que no hubo una entrega física del producto. Esta operación se llevó a cabo con la autorización previa del Gerente de Ventas.

Un caso similar se presentó con la salida registrada el día domingo 30 de junio. En este caso el depósito se efectuó ese mismo día y la facturación se realizó el lunes 1º de julio, debido al cierre del mes.

Para evitar este tipo de registros erróneos, se instruirá al personal encargado de facturación para que no asignen fechas de salida en días no laborables. Asimismo, se solicitará al proveedor del sistema que este tipo de movimiento no genere automáticamente una salida del almacén, quedando pendiente hasta la emisión de la guía de remisión.

b. Estacionalidad

En la figura 4, se presenta que los períodos de mayor venta son en mayo y octubre 2024, aunque los meses de noviembre y diciembre 2024 presentan montos elevados.

Figura 4 - Análisis de estacionalidad.

Reporte_03_Estacionalidad X														
	CODIGO	MES_1	MES_2	MES_3	MES_4	MES_5	MES_6	MES_7	MES_8	MES_9	MES_10	MES_11	MES_12	TOTAL
1	041811/2200N	-16,238.16	-55,441.54	-13,349.60	0.00	-81,765.76	-35,542.72	-11,820.40	-13,613.84	-37,254.64	-50,342.91	-35,210.40	-71,169.67	-421,749.64
2	042411/2200N	-20,197.80	-28,311.26	-14,962.86	-40,580.16	-137,033.40	-16,924.80	-38,945.90	-40,409.85	-42,518.28	-25,117.66	-53,866.71	-34,473.99	-493,342.67
3	05P11/2A	-5,083.00	0.00	0.00	-20,159.08	-55,670.92	-40,922.00	-10,182.34	0.00	-41,288.00	-97,170.04	0.00	-10,348.00	-280,823.38

Si bien es cierto, en la etapa posterior a la pandemia, se buscaron mercados en el sector avícola, deportes y otros, el mayor porcentaje de las ventas de la empresa, aproximadamente el 67%, corresponde al sector de la pesca, tanto artesanal como industrial.

Por ese motivo, la estacionalidad de este sector influye en el nivel de ventas que alcanzaron un mayor nivel en los meses de mayo y octubre 2024.

A esto se agrega que también hubo una mayor demanda en el exterior por lo que se incrementaron las exportaciones.

En 2023, la producción pesquera en Perú registró una notable contracción del 19.75% en comparación con el año anterior. Esta disminución se atribuyó principalmente a las condiciones oceanográficas adversas asociadas al fenómeno de El Niño, que afectaron la disponibilidad de especies clave como la anchoveta. En diciembre de 2023, la extracción de anchoveta se redujo en un 84.28% respecto al mismo mes de 2022.

El año 2024 mostró una recuperación importante. En mayo, el Ministerio de la Producción reportó un crecimiento del 329.2% en el valor bruto de producción del sector pesca extractivo en comparación con mayo de 2023. Este repunte se debió principalmente al incremento en la extracción de anchoveta para consumo humano indirecto, gracias a una exitosa primera temporada de pesca.

La segunda temporada de pesca de anchoveta en la región norte-centro, que inició el 1 de noviembre de 2024, concluyó el 21 de enero de 2025 y tuvo un resultado que fue el mejor desempeño registrado desde 2020.

c. Comprobantes de almacén duplicados

Como política del negocio se tiene establecido que los ingresos a almacén deben de realizar con un comprobante. En la figura 5, se presenta un mismo número de comprobante que sustenta diferentes ingresos al almacén en el mismo día. El monto de ingreso supera los S/25,000.

Figura 5 - Comprobantes de almacén duplicados.

CODIGO	NCODIGO	ITEM	COMPRBTE	FECHA	ALMACEN	CANT_INGR	CANT_SALI	SALDO	VAL_UNIT	VAL_TOTAL
1 041811/2200N	Red nylon 210/18 x 1 1/2" x 200 M.A., negro - (KG)	1	106-000185	04/01/2024	ALMA01	285.00	0.00	5.393.80	20.4200	5.819.70000
2 041811/2200N	Red nylon 210/18 x 1 1/2" x 200 M.A., negro - (KG)	2	106-000185	04/01/2024	ALMA01	285.00	0.00	5.678.80	20.4200	5.819.70000
3 042411/2200N	Red nylon 210/24 x 1 1/2" x 200 M.A., negro - (KG)	1	106-000185	04/01/2024	ALMA01	162.00	0.00	2.990.50	20.5800	3.333.96000
4 042411/2200N	Red nylon 210/24 x 1 1/2" x 200 M.A., negro - (KG)	2	106-000185	04/01/2024	ALMA01	328.00	0.00	3.318.50	20.5800	6.750.24000
5 05P11/2A	Cabo poliester torcido 1 1/2", alquitranado - (KG)	1	106-000185	04/01/2024	ALMA01	260.00	0.00	629.20	19.6000	5.096.00000

En efecto, se observa que el día 4 de enero de 2024 se registraron dos ingresos de Red Nylon 210/18 x 1 1/2" x 200 M.A., color negro, por 285 kg cada uno.

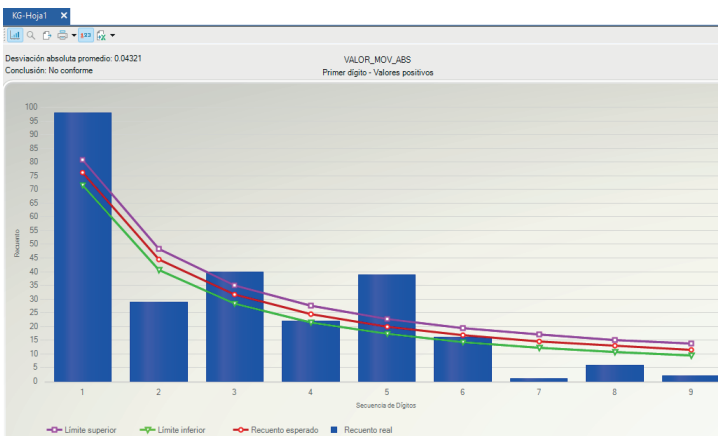
Ese mismo día, en la nota de ingreso 106-000185, también se registraron dos ingresos de Red nylon 210/24 210/24 x 1 1/2" x 200 M.A., color negro, con pesos de 162 Kg y 328 Kg respectivamente, además de un ingreso adicional de 260 Kg de Cabo poliéster torcido 1 1/2" de diámetro, alquitranado.

En el caso de las redes nylon 210/18, no se trata de una duplicidad en el registro, sino del ingreso de dos unidades correspondientes a pedidos de clientes distintos, las cuales fueron ingresadas al sistema por separado para un mejor control.

d. Análisis Benford

Las salidas de almacén han sido evaluadas mediante la regla de anti-fraude, con la utilización del algoritmo Benford. Al respecto, conviene precisar que la prueba Benford sirve para detectar posible manipulación de datos que podrían configurar fraude.

Figura 6 - Análisis Benford.



Las operaciones no se comportan según indica la teoría del primer dígito Benford, según se muestra en la figura 6. Esto podría deberse a ciertas reglas de negocio que no se han mapeado en esta fase inicial de análisis pero que convendría hacerse en fases posteriores.

Análisis de costos unitarios

Se ha analizado la información de inventarios de los tres principales productos, según su movimiento. Se ha logrado identificar los costos unitarios máximos y mínimos por cada ítem de interés. Los cuales presentan patrones de comportamiento adecuado en relación con el promedio calculado, según se muestra en la figura 7.

Figura 7 - Análisis de costos unitarios.

Reporte_09_Cunit		CODIGO	NUM_DE_REGS1	C_UNIT_SUMA	C_UNIT_MÁX	C_UNIT_MÍN	C_UNIT_PROMEDIO	C_UNIT_VARIANZA	C_UNIT_DES_STD
1	041811/2200N		101	2,058.5400	20.4200	20.3400	20.3816	0.0006	0.0238
2	05P11/2A		33	1,047.3040	19.9000	19.5200	19.7605	0.0104	0.1018
3	042411/2200N		99	2,045.6200	20.7600	20.5800	20.6628	0.0019	0.0432

Se ha verificado que los precios máximos y mínimos se encuentran dentro del rango promedio. Esta tabla sirve como insumo para verificaciones posteriores que lo comparen con los valores de venta en los comprobantes de pago emitidos.

De esa forma, se podrá tener pistas de la forma cómo se han estado vendiendo los inventarios, podría darse el caso que el costo según Kardex sea 20, pero se ha estado vendiendo a 12, es decir, por debajo del costo.

DISCUSIÓN

Los resultados de este estudio confirman que la implementación de herramientas de inteligencia artificial y automatización robótica en la contabilidad gerencial mejora significativamente la eficiencia del control financiero y la gestión de inventarios (CARRANZA-ALBERCA *et al.*, 2024; GLORIA ROBEL *et al.*, 2024). La detección de inconsistencias en el registro de comprobantes, movimientos en días no laborables y variaciones en costos unitarios respalda la necesidad de optimizar los procesos administrativos mediante tecnología avanzada. Estos hallazgos coinciden con estudios previos, como los de AL NATOUR *et al.* 2023; CHAMBI-AVALOS *et al.* (2025), quienes destacan que la analítica

avanzada y la automatización reducen los errores contables y fortalecen la toma de decisiones estratégicas en entornos empresariales dinámicos. No obstante, a diferencia de estos estudios, la presente investigación aborda específicamente el sector pesquero, demostrando que las fluctuaciones en las capturas y la estacionalidad de la producción requieren un enfoque diferenciado en la gestión financiera.

Asimismo, la aplicación del software IDEA permitió realizar un análisis más detallado de los estados financieros, identificando patrones clave en el movimiento de inventarios y en la trazabilidad de las operaciones. Este resultado refuerza las conclusiones de (CHÁVEZ-DÍAZ *et al.*, 2024; SETYAWAN, 2020) sobre la evolución de la contabilidad gerencial hacia un enfoque más estratégico, basado en el uso de big data y algoritmos predictivos. Sin embargo, a diferencia de investigaciones anteriores que se centran en grandes corporaciones, este estudio demuestra que la digitalización también es viable y beneficiosa para empresas del sector pesquero con estructuras organizativas más flexibles. Además, la detección de duplicidades en comprobantes de almacén y la prueba de Benford sugiere que la automatización contable no solo optimiza la gestión financiera, sino que también contribuye a la prevención del fraude y al fortalecimiento del control interno. Estos resultados subrayan la necesidad de seguir explorando la integración de modelos predictivos en la contabilidad gerencial para mejorar la competitividad del sector pesquero en un entorno económico cada vez más volátil.

CONCLUSIONES

La investigación demuestra la importancia de la contabilidad gerencial y el uso de herramientas tecnológicas avanzadas en la gestión financiera y operativa del sector pesquero peruano. A través del análisis de estados financieros y del rubro de inventarios, se identificaron patrones relevantes que afectan la eficiencia de la empresa estudiada, tales como inconsistencias en registros de almacén, movimientos en días no laborables y variaciones significativas en costos unitarios. Estos hallazgos evidencian la necesidad de mejorar los procesos de control y gestión financiera para optimizar el desempeño empresarial.

La implementación de tecnologías como el software IDEA ha permitido automatizar el análisis de información contable y financiera, facilitando la identificación de riesgos y oportunidades de mejora. Se confirma que la digitalización y el uso de algoritmos robóticos no solo incrementan la precisión en la evaluación de los datos, sino que también contribuyen a la toma de decisiones estratégicas basadas en información fiable y en tiempo real. Este enfoque innovador fortalece la capacidad de la empresa para adaptarse a cambios en el mercado y mejorar su rentabilidad.

Desde el punto de vista del control de inventarios, los resultados indican que aún existen desafíos en la gestión eficiente de las existencias, ya que se detectaron salidas de almacén en fechas no laborables y registros duplicados de comprobantes. Estos factores pueden afectar la trazabilidad de los productos y generar pérdidas financieras innecesarias. Por lo tanto, es fundamental que la empresa refuerce sus políticas de auditoría interna y adopte prácticas más estrictas para garantizar un manejo adecuado de sus recursos.

Finalmente, este estudio sugiere que la integración de modelos predictivos y herramientas de inteligencia artificial en la contabilidad gerencial podría representar el siguiente paso en la optimización del sector pesquero. La automatización y digitalización de los procesos contables no solo mejoran la eficiencia y reducen el margen de error, sino que también proporcionan ventajas competitivas sostenibles en un entorno de alta volatilidad. Se recomienda seguir explorando estas innovaciones para fortalecer la solidez financiera y operativa de las empresas en el sector.

REFERENCIAS

ABDULLAH, N. H. N.; SANUSI, S.; SAVITRI, E. The Role and Implications of Big Data on Strategic Management Accounting Practices: A Case Study in a Malaysian Manufacturing Company. **Management and Accounting Review**, v. 21, n. 1, p. 41–60, 2022.

ADAMOVA, G. A.; KHABIB, M. D.; TEPLYAKOVA, M. Y. The Problems with Information Support of Strategic Management. Em: **Studies in Systems, Decision and Control**. [s.l.: s.n.]. v. 282p. 33–41.

AL NATOUR, A. R. *et al.* The role of forensic accounting skills in fraud detection and the moderating effect of CAATs application: evidence from Egypt. **Journal of Financial Reporting and Accounting**, 12 set. 2023.

ALESANI, D. Financial management – budgeting, reporting, and leveraging financial data. Em: **Management of International Institutions and NGOs**. London: Routledge, 2023. p. 306–349.

ALSHARARI, N. M. Management Accounting Systems and Data Analytics in the United States: Evidence from Apple. Em: **Studies in Big Data**. [s.l: s.n.]. v. 147p. 141–152.

ARANDA, M. Evolution and state of the art of fishing capacity management in Peru: The case of the anchoveta fishery | Evolución y situación de la gestión de la capacidad de pesca en Perú: El caso de la pesquería de anchoveta. **Pan-American Journal of Aquatic Sciences**, v. 4, n. 2, p. 146–153, 2009.

ARIAS SCHREIBER, M.; ÑIQUEN, M.; BOUCHON, M. Coping strategies to deal with environmental variability and extreme climatic events the Peruvian anchovy fishery. **Sustainability**, v. 3, n. 6, p. 823–846, 2011.

CALOTĂ, T. O. Managerial Accounting. Em: **CSR and Management Accounting Challenges in a Time of Global Crises**. [s.l: s.n.]. p. 233–246.

CARRANZA-ALBERCA, C. A. *et al.* Aplicación de CAAT IDEA en el Análisis Tributario del IGV: Caso Empresa de Servicios. Em: **Tecnologias e Metodologias Ativas: a interdisciplinaridade tecnológica em pesquisa - Volume 5**. São Paulo: Editora Científica Digital, 2024. v. 5p. 193–212.

CHAMBI-AVALOS, M. L. A. *et al.* Aplicación de CAATs en la Auditoría Tributaria Preventiva: Libro Diario - Gastos. Em: **Desafíos e Oportunidades na Educação Contemporânea: inclusão, tecnologia e políticas públicas**. [s.l.] Editora Científica Digital, 2025. p. 87–104.

CHÁVEZ-DÍAZ, J. M. Evolución de las Computer Assisted Audit Techniques en el proceso de Fiscalización Tributaria, 2024. **Cátedra Villarreal Posgrado**, v. 3, n. 1, p. 21–30, 4 jul. 2024.

CHÁVEZ-DÍAZ, J. M. *et al.* Artificial intelligence in accounting and auditing: bibliometric analysis in Scopus 2020–2023. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 36, n. 2, p. 1319, 1 nov. 2024.

CHIPRIYANOVA, G.; KRASTEVA-HRISTOVA, R. Technological aspects of Accounting Automation System as a Decision Support System. **ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference**, v. 2, p. 28–33, 13 jun. 2023.

COAYLA, E. *et al.* Industrial Marine Fishing in the Face of Climate Change in Peru. **Journal of Marine Sciences**, v. 2023, 2023.

DIAKOV, V. V. Evaluation of variations in production costs basing on their operational analysis in managerial accounting. **Actual Problems of Economics**, n. 2, p. 229–233, 2010.

GALARZA, E.; KÁMICHE, J. **Pesca artesanal: oportunidades para el desarrollo regional**. 1. ed. Lima: Universidad del Pacífico, 2015.

GLORIA ROBEL, E. *et al.* **Utilization of CAAT in Continuous Auditing: A Case Applied to the Ministry of Health, Peru 2022-2023**. Proceedings of the 22nd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI 2024): “Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research. **Anais...Latin American and Caribbean Consortium of Engineering Institutions**, 2024. Disponível em: <<https://laccei.org/LACCEI2024-CostaRica/meta/FP1329.html>>

GOZZER-WUEST, R.; ALONSO-POBLACIÓN, E.; TINGLEY, G. A. Identifying priority areas for improvement in Peruvian Fisheries. **Marine Policy**, v. 129, 2021.

HEIDARI, I. Management accounting and its role in organizations. **Life Science Journal**, v. 9, n. 4, p. 545–550, 2012.

HENG, M. S. H.; FIEDLER, B.; KANDUNIAS, C. **Implications of supply chain management for accounting information systems**. 9th Pacific Asia Conference on Information Systems: I.T. and Value Creation, PACIS 2005. **Anais...**2005.

JACKOVÁ, A. Use of management accounting in business management. Em: **Production Management and Engineering Sciences**. [s.l.] CRC Press, 2015. p. 113–118.

JAMES, M. A. *et al.* To ignore or mitigate – Economic implications of an illegal artisanal trawl fishery in northern Peru. **Marine Policy**, v. 158, 2023.

LUEG, R.; ILIEVA, D. Customer Profitability Analysis in decision-making–The roles of customer characteristics, cost structures, and strategizing. **PLOS ONE**, v. 19, n. 5, p. e0296974, 22 maio 2024.

MALDONADO-CUEVA, P. D.; CRIADO-DÁVILA, Y. V.; FERNANDEZ-BEDOYA, V. H. **Foreign investment in the fishing sector after the Free Trade Agreement between Peru and China, 2009-2021 | La inversión extranjera en el sector pesquero después del Acuerdo de Libre Comercio entre Perú y China, 2009-2021**. Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology. **Anais...**2024.

MICULESCU, C. Managerial accounting and decision-making in terms of sustainable development. **Quality - Access to Success**, v. 14, n. SUPPL.2, p. 14–20, 2013.

ONCIOIU, I. **CSR and Management Accounting Challenges in a Time of Global Crises**. [s.l.] IGI Global, 2021.

PICKERING, M. An Evolution: Turning Management Accounting into a Strategic Function. Em: **Effective Implementation of Transformation Strategies**. Singapore: Springer Nature Singapore, 2022. p. 177–201.

PORTER, M. **Competitive Advantage Creating and Sustaining Superior Performance**. New York: The Free Press, 1985.

SETYAWAN, E. C. The use of Benford´s Law in performance Audit to Detect Fraud in the State Expenditure Transactions Assisted by IDEA Software (A Case Study at the Transportation Ministry Work Units). **Asia Pacific Fraud Journal**, v. 5, n. 1, p. 147, 30 jun. 2020.

SUÁREZ-CASTAÑEDA, M. I. Environmental Balance and Species Protection in Peru: The Selachimorpha Shark as a Fundamental Element for the Preservation of Life and Ecosystems | Equilibrio ambiental y protección de las especies en Perú: El tiburón Selachimorpha como elemento fundamen. **Clio. Revista de Historia, Ciencias Humanas y Pensamiento Crítico**, v. 5, n. 9, p. 246–266, 2025.

TICSE-VILLANUEVA, E. *et al.* **"Importance of the Fishing Industry in Peru, an approach towards its sustainable development"**. Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology. **Anais...**2021.

VLADIMIROVNA GALAUDINOVA, V. Automated Strategic Management Batch Accounting in a Processing Industry. **TEM Journal**, v. 8, n. 3, p. 1025–1030, 28 ago. 2019.

Anexo 1 – Uso de IA

Se utilizó el ChatGPT para generar los ratios financieros.

Procedimiento:

Se subió al chatbot los estados financieros en formato Excel.

Se utilizó el siguiente Prompt, para la generación de ratios:

"A partir del archivo, genera los ratios más importantes bajo el enfoque de contabilidad gerencial. Preséntalo en una tabla, en una columna debe estar indicada la formula, en otra los valores aritméticos y en otra el valor final calculado".

Se verificó el resultado final.

Anexo 2 – Algoritmo robótico

Repositorio: https://github.com/jorgechavez6816/managerial_accounting

APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA OCEANOGRAFIA: INOVAÇÕES, DESAFIOS E PERSPECTIVAS

Erick Paiva Barbieri
Universidade Federal Tecnológica do Paraná (UTFPR)

Edison Barbieri
Instituto de Pesca-Governo do Estado de São Paulo

RESUMO

Inteligência Artificial (IA) representa uma das mais profundas transformações tecnológicas da era contemporânea, promovendo avanços disruptivos em diversas áreas do conhecimento. No contexto da oceanografia — ciência multidisciplinar que estuda os processos físicos, químicos, biológicos e geológicos dos oceanos — a aplicação da IA tem permitido novas abordagens para a coleta, análise e interpretação de dados em escala global. Este capítulo explora as aplicações emergentes da IA na oceanografia, articulando os principais eixos tecnológicos que estruturam a Quarta Revolução Industrial e suas implicações para os sistemas naturais e sociais.

Palavras-chave: inteligência; artificial; oceanografia; aplicação.

Algoritmos de Aprendizado de Máquina e Deep Learning na Oceanografia

O uso de algoritmos de *machine learning* (ML) e *deep learning* (DL) tem revolucionado profundamente a análise de dados oceanográficos, especialmente diante da crescente disponibilidade de grandes volumes de dados provenientes de satélites de observação terrestre, boias de monitoramento, veículos autônomos subaquáticos, sensores in situ e modelos numéricos de previsão (Zhao *et al* 2024). Tais algoritmos têm demonstrado alto desempenho na modelagem de fenômenos oceânicos complexos, superando em muitos casos as abordagens estatísticas tradicionais em termos de acurácia, escalabilidade e capacidade de generalização.

Modelos supervisionados, como *Support Vector Machines* (SVM), *Random Forests* e redes neurais artificiais (ANNs), e modelos não supervisionados, como algoritmos de agrupamento (*clustering*) e análise de componentes principais (PCA), têm sido empregados com sucesso em diversas frentes, tais como (Zhao *et al.*, 2024):

- Previsão de padrões de circulação oceânica e temperatura da superfície do mar (SST): Utilizando séries temporais multivariadas e modelos de redes recorrentes (RNNs) ou redes de memória de longo curto prazo (LSTM), pesquisadores têm conseguido prever com maior antecedência eventos como El Niño e La Niña, bem como anomalias térmicas associadas à acidificação dos oceanos e ao branqueamento de corais (Gray *et al.*, 2024)
- Classificação automatizada de habitats bentônicos e substratos marinhos: Com o uso de imagens multiespectrais e dados batimétricos, algoritmos de visão computacional, em especial as redes convolucionais (CNNs), têm permitido a classificação de habitats como recifes, pradarias de fanerógamas e fundos sedimentares com grande precisão, contribuindo diretamente para o planejamento espacial marinho e conservação de ecossistemas vulneráveis (Li *et al.* 2020).
- Detecção de organismos marinhos, fitoplâncton e detritos plásticos: Técnicas de detecção de objetos e segmentação semântica, baseadas em redes como YOLO (*You Only Look Once*) e U-Net, são empregadas para identificar padrões visuais característicos de espécies biológicas ou resíduos sólidos flutuantes em imagens captadas por drones e

satélites (Bachimanch et al. 2024). Esse tipo de aplicação é particularmente relevante para o monitoramento de biodiversidade e avaliação da poluição por micro e macroplásticos.

- Modelagem de processos biogeoquímicos: Algoritmos de ML têm sido utilizados para estimar concentrações de clorofila-a, oxigênio dissolvido e carbono orgânico particulado a partir de dados de sensoriamento remoto e medições in situ (Sadaïappan et al. 2023). Esses modelos contribuem para o entendimento da produtividade primária e da dinâmica do carbono nos oceanos, o que é crucial em cenários de mudanças climáticas.
- Análise de padrões acústicos para estudos de bioacústica marinha: Técnicas de DL são empregadas na classificação de vocalizações de cetáceos, peixes e invertebrados, bem como na detecção de ruído antrópico, permitindo inferir aspectos ecológicos e comportamentais das espécies marinhas e avaliar o impacto da poluição sonora nos ecossistemas aquáticos.

A principal vantagem dessas técnicas reside na sua capacidade de lidar com conjuntos de dados altamente complexos, não-lineares, heterogêneos e com múltiplas dimensões — características que refletem a própria natureza dinâmica e interconectada dos sistemas oceânicos. Enquanto modelos estatísticos clássicos exigem fortes suposições sobre a distribuição dos dados e relações lineares entre variáveis, os modelos baseados em ML e DL são mais flexíveis e adaptativos, aprendendo diretamente a partir dos dados.

A aplicação desses algoritmos demanda cautela científica e metodológica. Modelos de *deep learning*, notadamente redes neurais profundas com múltiplas camadas, são frequentemente considerados modelos de caixa-preta (*black-box models*), pois, embora apresentem alta acurácia preditiva, tendem a dificultar a interpretação causal e a rastreabilidade dos resultados. Essa limitação é particularmente crítica em contextos de gestão ambiental e formulação de políticas públicas, onde decisões precisam ser justificadas de forma transparente.

Nesse sentido, é imprescindível o uso de técnicas de interpretabilidade, como o *SHAP* (SHapley Additive exPlanations) e o *LIME* (Local Interpretable Model-agnostic Explanations), que buscam elucidar as relações entre variáveis de entrada e as saídas dos modelos, oferecendo maior clareza sobre os fatores

determinantes nas previsões (Hassan *et al.*, 2025). Além disso, o emprego de validação cruzada, testes independentes e abordagens de *ensemble learning* contribuem significativamente para assegurar a robustez, a generalização e a confiabilidade dos modelos desenvolvidos.

Deve-se observar que a eficácia dos algoritmos depende diretamente da qualidade e da representatividade dos dados utilizados em seu treinamento. Dados ruidosos, incompletos ou enviesados podem comprometer o desempenho do modelo e levar a conclusões equivocadas, com implicações negativas para a ciência e a gestão marinha. Por essa razão, os esforços de padronização e curadoria de dados oceanográficos, bem como a construção de repositórios abertos e bem documentados, são fundamentais para o avanço da aplicação ética e eficaz da IA na oceanografia.

O uso de *machine learning* e *deep learning* representa um novo paradigma na análise dos oceanos — um paradigma orientado por dados, mas que demanda forte embasamento teórico, responsabilidade ética e integração interdisciplinar para alcançar seu pleno potencial transformador.

IA na Indústria 4.0 e a Oceanografia Operacional

A convergência entre a IA e os princípios da Indústria 4.0 tem impulsionado uma revolução tecnológica nos sistemas de monitoramento, previsão e gestão dos oceanos. Essa nova era — marcada pela digitalização, automação inteligente e conectividade em tempo real — tem promovido avanços sem precedentes na capacidade de observação, modelagem e resposta às dinâmicas oceânicas, especialmente no campo da oceanografia operacional.

A Indústria 4.0, baseada na integração de tecnologias como Internet das Coisas (IoT), computação em nuvem, *big data analytics*, robótica avançada e IA, permite a criação de sistemas ciberfísicos interconectados que operam com autonomia e inteligência distribuída. No contexto oceanográfico, essa abordagem tem viabilizado o desenvolvimento de uma nova geração de dispositivos e infraestruturas inteligentes, tais como (Khan *et al.*, 2025):

- Veículos Autônomos Subaquáticos (AUVs) e Veículos de Superfície Não Tripulados (USVs), equipados com sensores acústicos, câmeras multiespectrais, sondas de CTD e sistemas de navegação autônoma,

que são capazes de operar por longos períodos em ambientes hostis e remotos. Algoritmos de IA são utilizados para processamento em tempo real, tomada de decisão adaptativa e otimização de trajetórias de coleta de dados.

- Robôs biomiméticos e drones marinhos, que se inspiram em formas e padrões de movimentação de organismos aquáticos, sendo capazes de realizar tarefas de inspeção de infraestruturas submarinas, mapeamento de habitats, monitoramento de recifes de corais e detecção de fontes de poluição difusa.
- Plataformas móveis e flutuantes inteligentes, como boias inteligentes e *gliders*, que transmitem dados em tempo real via satélite ou redes móveis, permitindo o monitoramento contínuo de variáveis físico-químicas e biológicas em regiões estratégicas, como zonas costeiras, áreas de surgência, ou rotas de navegação comercial.

No escopo da oceanografia operacional, que visa fornecer informações tempestivas, precisas e acessíveis para apoiar atividades humanas no oceano — como navegação, pesca, exploração offshore e gestão de riscos naturais — a IA tem se mostrado uma ferramenta transformadora. Os sistemas inteligentes de previsão e apoio à decisão desenvolvidos com base em IA permitem (Ardanza *et al.*, 2019):

- Otimização de rotas de embarcações comerciais e de pesca, reduzindo custos operacionais, consumo de combustível e emissões de gases de efeito estufa. Redes neurais e algoritmos de reforço são usados para aprender padrões de correntes, ventos e ondas, oferecendo rotas mais seguras e eficientes.
- Previsão de condições oceanográficas e meteorológicas com alta acurácia, integrando dados observacionais e modelos numéricos com técnicas de aprendizado profundo (como LSTM e *convolutional neural networks*), aumentando a confiabilidade das previsões de marés, ondas, ressacas, tsunamis e derrames de óleo.
- Implementação de sistemas de alerta precoce para desastres naturais, como inundações costeiras, erosão marinha, blooms de algas nocivas e ciclones tropicais. Tais sistemas aprendem com dados históricos e em tempo real, e são capazes de detectar anomalias e emitir alertas

com antecedência suficiente para mitigar danos socioambientais.

- Gestão adaptativa da zona costeira, combinando sensoriamento remoto, dados geoespaciais e IA para identificar áreas vulneráveis à subida do nível do mar, erosão, salinização de aquíferos e impactos de eventos extremos, fornecendo subsídios técnicos à tomada de decisão por gestores públicos.

Além disso, a integração de IA com sistemas de digital twins (gêmeos digitais) de ambientes marinhos está emergindo como uma nova fronteira da oceanografia operacional. Esses modelos virtuais dinâmicos, alimentados por fluxos contínuos de dados empíricos e simulados, permitem simular cenários futuros, testar políticas de uso do espaço marítimo e responder rapidamente a mudanças no ambiente.

O avanço dessa “oceanografia automatizada” demanda uma infraestrutura robusta de comunicação e interoperabilidade de dados, além de protocolos éticos e regulamentares que garantam a segurança, a transparência e o uso responsável das tecnologias. A crescente autonomia dos sistemas ciberfísicos levanta também questões sobre a soberania digital e o acesso equitativo às inovações tecnológicas, especialmente em regiões costeiras vulneráveis e em países em desenvolvimento.

Do ponto de vista científico, a integração entre oceanografia e Indústria 4.0 oferece oportunidades únicas para estudos multidisciplinares, promovendo uma oceanografia mais responsiva, inclusiva e orientada a soluções. O desenvolvimento de novas habilidades profissionais, como a ciência de dados aplicada ao oceano (*ocean data science*), torna-se uma demanda estratégica para formar especialistas capazes de interpretar os dados produzidos por essas tecnologias e transformá-los em conhecimento útil para a sociedade.

A aplicação da IA no contexto da Indústria 4.0 está remodelando o modo como conhecemos e interagimos com o ambiente marinho, alavancando uma nova era de governança inteligente dos oceanos, baseada em automação, previsão e adaptabilidade contínua.

Saúde Ambiental Marinha: Convergência com a Medicina e a Saúde Pública

A aplicação de algoritmos de IA na área da saúde tem evoluído rapidamente, especialmente no campo da análise de grandes volumes de dados clínicos e ambientais, detecção de doenças e modelagem de riscos sanitários. Na oceanografia, essa expertise tem sido transposta para o monitoramento da saúde ambiental marinha, permitindo (Albahra *et al.*, 2023):

- Detecção precoce de riscos ecotoxicológicos, por meio da análise de séries temporais de concentração de contaminantes (metais pesados, hidrocarbonetos, microplásticos) e da modelagem de padrões de bioacumulação em organismos marinhos bioindicadores.
- Avaliação do impacto de poluentes emergentes como nanopartículas, fármacos e compostos hormonais, com o uso de algoritmos de *machine learning* para identificar correlações entre exposições químicas e alterações fisiológicas ou genéticas em peixes, moluscos e aves marinhas.
- Predição de blooms de algas nocivas (HABs), integrando dados físico-químicos, meteorológicos e satelitais com redes neurais profundas, o que permite alertas antecipados com maior acurácia e minimização de impactos à saúde humana e à aquicultura.

Essas aplicações representam uma fusão entre a ecotoxicologia, a saúde pública e a biotecnologia ambiental, inaugurando uma nova abordagem preditiva e preventiva para a gestão da saúde dos ecossistemas costeiros e marinhos.

Economia Azul, Finanças e Gestão de Riscos Marítimos

A interseção entre IA e finanças tem originado uma gama de ferramentas analíticas avançadas — desde modelos de crédito até análise de riscos sistêmicos — que podem ser recontextualizadas no âmbito da economia azul, ou seja, a economia baseada no uso sustentável dos recursos oceânicos.

- Modelos preditivos de avaliação de risco já vêm sendo empregados por seguradoras marítimas e empresas de exploração offshore para estimar a probabilidade de eventos catastróficos, como tempestades, acidentes com embarcações ou falhas estruturais em plataformas. A IA permite incorporar variáveis oceanográficas dinâmicas nesses

cálculos, como mudanças sazonais de correntes ou intensificação de ciclones tropicais devido às mudanças climáticas.

- Análise financeira de projetos de energia renovável marinha (eólica offshore, energia das ondas e marés), considerando a variabilidade natural do ambiente e suas implicações no retorno do investimento. Algoritmos evolutivos e aprendizado por reforço auxiliam na simulação de cenários de longo prazo para viabilização econômica desses empreendimentos.
- Gestão inteligente da pesca e da aquicultura, com sistemas de apoio à decisão financeira que integram previsões oceanográficas, dados de mercado e estimativas de produtividade biológica para otimizar cadeias produtivas e minimizar perdas econômicas associadas a eventos ambientais adversos.

Esses exemplos demonstram que o uso transversal da IA nos domínios econômico-financeiros permite construir uma base sólida para o desenvolvimento sustentável dos oceanos, conectando rentabilidade com responsabilidade ecológica.

Educação em Ciências Marinhas: Inclusão Digital e Democratização do Conhecimento

No setor educacional, a IA tem desempenhado um papel central na personalização do ensino, na criação de plataformas adaptativas e na inclusão digital em diferentes contextos sociais. Tais inovações também têm encontrado espaço no ensino e na popularização das ciências marinhas, particularmente em regiões remotas ou com acesso limitado à educação formal.

- Tutores inteligentes e sistemas de aprendizagem adaptativa oferecem conteúdo em oceanografia adaptado ao ritmo e ao nível de conhecimento de cada aluno, promovendo maior engajamento e retenção do aprendizado. Esses ambientes digitais são especialmente eficazes no ensino à distância de disciplinas complexas, como modelagem oceânica, climatologia e biogeoquímica marinha.
- Ambientes de realidade aumentada e virtual integrados com IA, que simulam ecossistemas marinhos em tempo real, têm se mostrado eficazes tanto no ensino básico quanto na formação de especialistas.

Estudantes podem explorar recifes de coral, acompanhar migrações de espécies ou modelar cenários de acidificação dos oceanos em uma experiência interativa e imersiva.

- Sistemas multilíngues com Processamento de Linguagem Natural (NLP) viabilizam o acesso ao conhecimento oceanográfico para comunidades costeiras diversas, superando barreiras linguísticas e promovendo a cidadania científica.

Essas tecnologias educacionais representam não apenas um avanço pedagógico, mas uma ferramenta poderosa de justiça cognitiva, permitindo a inclusão de comunidades tradicionalmente excluídas da construção do saber oceanográfico, como povos indígenas, comunidades ribeirinhas e pescadores artesanais.

Chatbots e Assistentes Virtuais na Ciência Oceânica: Caminhos para uma Interação Inteligente com a Oceanografia

O uso de chatbots e assistentes virtuais baseados em IA na ciência oceânica representa um campo emergente, ainda incipiente em termos de implementação prática, mas repleto de potencial transformador. Esses sistemas interativos, que operam com base em técnicas de *Natural Language Processing* (NLP), aprendizado de máquina e análise semântica, têm se consolidado em setores como o comércio eletrônico, o atendimento ao cliente e a educação. Na oceanografia, sua introdução propõe novas formas de interação entre humanos e dados marinhos, favorecendo a democratização do conhecimento, a eficiência na gestão de informações e a inovação em serviços de monitoramento ambiental.

Aplicações Promissoras na Oceanografia

O desenvolvimento de chatbots especializados em oceanografia pode viabilizar soluções aplicadas em múltiplas frentes (Pergantis *et al.*, 2025):

- Acesso público a dados marinhos em tempo real: chatbots integrados a sistemas de sensoriamento remoto e plataformas oceanográficas (como boias, AUVs ou estações costeiras) podem fornecer informações atualizadas sobre marés, correntes, temperatura da água, índice UV e previsão de ressacas a públicos diversos, como pescadores artesanais,

navegadores recreativos, gestores ambientais e comunidades costeiras.

- Apoio à gestão e governança costeira: assistentes virtuais podem atuar como intermediários entre bancos de dados técnicos e gestores públicos, oferecendo relatórios automatizados, recomendações baseadas em modelos preditivos e alertas sobre riscos ambientais (como erosão costeira, contaminação marinha ou espécies invasoras).
- Educação em oceanografia e alfabetização oceânica: ambientes virtuais de aprendizagem podem incorporar chatbots capazes de responder a perguntas, sugerir leituras complementares, adaptar conteúdos ao nível de conhecimento dos estudantes e promover simulações interativas. Isso pode ser particularmente eficaz no ensino básico, na formação técnica e na capacitação de professores em áreas carentes de infraestrutura educacional.
- Apoio a pesquisadores em atividades de campo e laboratório: sistemas de assistentes virtuais podem automatizar tarefas como o preenchimento de fichas de coleta, a checagem de metadados, o envio de lembretes sobre calibração de equipamentos, a organização de cronogramas de amostragem e até a sugestão de protocolos baseados em normas técnicas previamente indexadas.
- Turismo científico e ecoturismo marinho: chatbots podem atuar como guias digitais em áreas costeiras e unidades de conservação, fornecendo informações sobre espécies locais, regulamentações de visitação, condições ambientais em tempo real e conteúdo educativo sobre os ecossistemas marinhos.

Desafios Técnicos e Epistemológicos

Apesar de seu potencial, a adoção de chatbots na ciência oceânica enfrenta uma série de obstáculos que demandam abordagens críticas e investimentos em pesquisa aplicada (Kooli *et al.*, 2023):

- Compreensão contextual e ambiguidade linguística: ambientes marinhos envolvem terminologias técnicas e contextos complexos que dificultam a interpretação precisa por parte de assistentes baseados em NLP. A ambiguidade de termos (ex.: “corrente” pode se referir a fluxo de água ou tendência de dados) exige refinamento semântico e treinamento com corpora específicos da área oceanográfica.

- Integração com bancos de dados heterogêneos: a ciência oceânica opera com dados provenientes de diversas fontes (modelos numéricos, sensores remotos, amostragem in situ), com formatos, escalas e graus de curadoria variados. A interoperabilidade entre essas bases e os sistemas de IA requer o uso de ontologias marinhas, metadados padronizados e protocolos abertos (como OGC, NetCDF e APIs RESTful).
- Privacidade, ética e governança algorítmica: é fundamental garantir que os dados utilizados e as interações com os usuários respeitem diretrizes éticas, especialmente em contextos sensíveis como populações tradicionais costeiras, dados de biodiversidade ou informações geopolíticas sobre zonas econômicas exclusivas. A transparência dos algoritmos e a auditabilidade das decisões automatizadas são elementos essenciais para assegurar a confiança pública.
- Barreiras de conectividade e infraestrutura: muitas regiões costeiras ou ilhas remotas carecem de acesso à internet de qualidade, o que limita a aplicação de soluções baseadas em nuvem. Estratégias que incorporem tecnologias offline-first ou edge computing são alternativas viáveis para ampliar a inclusão digital marinha.

Perspectivas Futuras e Interdisciplinaridade

O avanço de chatbots e assistentes virtuais na ciência oceânica dependerá de uma colaboração estreita entre cientistas marinhos, engenheiros de software, Analistas de sistemas, linguistas computacionais, designers de interface e especialistas em cognição humana. Além disso, a utilização dessas ferramentas pode fortalecer a ciência cidadã, ao permitir que não especialistas interajam com dados complexos de forma acessível e intuitiva, contribuindo com informações locais, validação de dados e sugestões para políticas públicas.

Iniciativas internacionais, como a Década da Ciência Oceânica para o Desenvolvimento Sustentável (2021–2030) da UNESCO, vêm promovendo a criação de infraestruturas digitais abertas e inclusivas. A implementação de chatbots orientados por esses princípios pode representar um passo decisivo para garantir que o oceano seja compreendido, valorizado e protegido por todos os segmentos da sociedade.

Ética e Responsabilidade: O Dilema da IA Oceânica

A crescente inserção da IA nos sistemas de monitoramento, previsão e gestão dos oceanos traz consigo não apenas avanços técnicos e operacionais, mas também uma série de desafios éticos e políticos que não podem ser negligenciados. A “IA oceânica” — como vem sendo chamada a aplicação de sistemas algorítmicos em contextos marinhos — impõe uma reflexão urgente sobre as formas como produzimos, interpretamos e utilizamos dados, sobretudo quando tais informações impactam comunidades humanas e ecossistemas marinhos frágeis.

Privacidade, Consentimento e Direitos Digitais nas Zonas Costeiras

A coleta massiva de dados ambientais por sensores, satélites, drones e veículos autônomos, quando combinada a metadados sobre uso do solo, pesca artesanal, turismo ou deslocamentos populacionais, pode representar uma ameaça à privacidade de comunidades costeiras. Em muitos casos, essas populações — como povos tradicionais, pescadores artesanais ou ribeirinhos — sequer são consultadas sobre o uso e destino dos dados extraídos de seus territórios ou modos de vida. Tal ausência de consentimento configura uma violação de direitos fundamentais, além de contrariar os princípios da Convenção sobre Diversidade Biológica (CDB) e da Declaração das Nações Unidas sobre os Direitos dos Povos Indígenas.

Mais ainda, o uso de IA em processos decisórios (por exemplo, no ordenamento de áreas marinhas protegidas ou na concessão de licenças para atividades econômicas no mar) pode ocorrer de maneira opaca, com algoritmos “caixa preta” cujos critérios não são compreendidos pelas comunidades afetadas. Isso fere o princípio da justiça procedimental, que exige transparência e participação nos processos de tomada de decisão ambiental.

Vieses Algorítmicos e Injustiça Ambiental

Outro risco ético importante reside nos vieses embutidos nos algoritmos. Modelos preditivos e classificadores utilizados na avaliação de riscos ecológicos, previsão de desastres naturais ou alocação de recursos de monitoramento podem ser treinados com bases de dados incompletas, historicamente enviesadas ou

desatualizadas. Isso pode levar à priorização indevida de determinadas regiões, espécies ou setores econômicos, em detrimento de áreas menos monitoradas ou populações marginalizadas.

Por exemplo, algoritmos que preveem áreas críticas de sobrepesca com base apenas em dados de frotas industriais, desconsiderando a pesca de subsistência ou comunitária, podem reforçar desigualdades no acesso aos recursos marinhos. Da mesma forma, sistemas de alerta baseados em redes neurais mal calibradas podem falhar ao prever eventos extremos em regiões tropicais, historicamente sub-representadas nos datasets globais.

Nesse sentido, a ética da IA oceânica deve incorporar os princípios da justiça ambiental, reconhecendo desigualdades históricas, sociais e geográficas na produção do conhecimento científico e nas decisões sobre o uso dos oceanos.

Governança Algorítmica e Responsabilidade Coletiva

Diante desses desafios, é imperativo que os projetos envolvendo IA na oceanografia sejam regidos por uma governança algorítmica robusta, que inclua:

- **Transparência:** os modelos e suas lógicas internas devem ser acessíveis e auditáveis por pesquisadores independentes, tomadores de decisão e comunidades impactadas;
- **Explicabilidade:** os resultados gerados pelos algoritmos precisam ser compreensíveis, não apenas para especialistas, mas também para usuários finais e públicos não técnicos;
- **Responsabilidade compartilhada:** cientistas, engenheiros de dados, formuladores de políticas públicas e usuários locais devem co-participar das etapas de planejamento, desenvolvimento, implementação e revisão dos sistemas de IA;
- **Revisões éticas contínuas:** projetos de longo prazo devem prever instâncias permanentes de avaliação ética, incluindo a possibilidade de corrigir rotas, atualizar critérios de decisão e incorporar novas vozes ao processo.

Esses elementos são particularmente relevantes no contexto da chamada economia azul, onde há grande pressão econômica e política sobre os recursos

marinhos, e onde os sistemas de IA podem ser instrumentalizados para legitimar decisões tecnocráticas e pouco democráticas.

Regulação Internacional e Cooperação Multinível

A ausência de normativas internacionais específicas para a IA aplicada ao ambiente marinho é um vácuo preocupante. Apesar de existirem marcos gerais — como a Recomendação da UNESCO sobre a Ética da Inteligência Artificial (2021) —, ainda não há tratados vinculantes que regulem o uso de algoritmos em águas internacionais ou que protejam os direitos informacionais de comunidades costeiras em diferentes jurisdições.

Assim, torna-se urgente promover uma governança transnacional da IA oceânica, que envolva organismos multilaterais (como a ONU, a IOC/ UNESCO, a FAO e o IMO), instituições de pesquisa, sociedade civil e o setor privado. É necessário garantir que os princípios de equidade, sustentabilidade e precaução estejam no centro da agenda tecnológica que guia a próxima geração de ferramentas digitais voltadas para o oceano.

Por uma Oceanografia Ética e Inclusiva

A aplicação ética da IA na oceanografia deve ir além de códigos de conduta ou declarações de boas intenções. Ela requer uma transformação estrutural nas formas como concebemos e aplicamos o conhecimento científico no mar. Isso implica reconhecer os oceanos não apenas como espaços de exploração ou observação, mas como ecossistemas interconectados com dimensões sociais, culturais e espirituais profundas.

A transição para uma “oceanografia algorítmica” não pode ocorrer à revelia das populações humanas e não humanas que dependem do oceano. Somente por meio da construção de sistemas transparentes, inclusivos e responsáveis, poderemos assegurar que os avanços da IA sirvam, de fato, ao bem comum planetário e não à reprodução de desigualdades e exclusões.

O Futuro do Trabalho na Oceanografia e a Automação

A emergência da IA e das tecnologias de automação está promovendo uma transformação profunda e estrutural nas práticas laborais da oceanografia

contemporânea. O que antes se caracterizava como uma ciência altamente dependente de observações empíricas diretas, de longas campanhas em alto-mar e da análise manual de amostras e registros ambientais, passa agora a se apoiar em sistemas automatizados, sensores inteligentes, veículos autônomos e plataformas computacionais de alta performance.

Do Campo ao Código: Uma Nova Configuração das Práticas Oceanográficas

A automação baseada em IA tem deslocado significativamente o eixo de atuação dos profissionais da oceanografia. Atividades repetitivas e intensivas em trabalho físico — como coleta manual de dados, triagem de organismos, ou planilhamento de séries temporais — estão sendo substituídas por tarefas mais sofisticadas e intelectualmente exigentes, como (Schwing, 2022):

- **Curadoria de dados e validação de séries ambientais:** requer conhecimento estatístico e domínio de padrões oceânicos, além da habilidade de lidar com grandes volumes de dados (Big Data);
- **Modelagem numérica e preditiva:** baseada em redes neurais, aprendizado profundo (deep learning) e assimilação de dados em modelos dinâmicos de circulação oceânica;
- **Desenvolvimento e aplicação de algoritmos:** para detecção automática de eventos oceanográficos (ex. upwellings, marés vermelhas), reconhecimento de padrões e previsão de riscos climáticos e ambientais;
- **Interpretação crítica dos outputs algorítmicos:** tarefa essencial, pois exige que o profissional compreenda não apenas os resultados, mas os limites, premissas e possíveis vieses dos modelos utilizados.

Essa nova configuração profissional exige uma profunda requalificação técnica e epistemológica dos oceanógrafos, que não podem mais se apoiar exclusivamente em uma formação tradicional baseada nas ciências naturais. É necessário incorporar um conjunto de novas competências, entre as quais se destacam:

- Ciência de dados aplicada ao oceano (Ocean Data Science);
- Programação em linguagens como Python, R, Julia e MATLAB;

- Pensamento computacional e lógica algorítmica;
- Visualização interativa e comunicação de dados complexos para públicos diversos.

A Interdisciplinaridade como Imperativo Formativo

A oceanografia está se tornando, cada vez mais, um campo híbrido, situado na confluência entre as geociências, a biologia marinha, as engenharias (particularmente a engenharia de automação e robótica), a ciência da computação e as ciências sociais aplicadas à governança marinha. Nesse sentido, a formação do novo profissional oceanógrafo deve ser repensada à luz de um currículo profundamente interdisciplinar, que vá além da integração superficial de saberes.

É crucial que os programas de graduação e pós-graduação em oceanografia passem a incluir disciplinas obrigatórias em IA, ética algorítmica, aprendizado de máquina, análise multivariada de grandes conjuntos de dados e simulações computacionais. Além disso, devem ser incentivadas práticas pedagógicas baseadas em projetos interdisciplinares, hackathons científicos e residências tecnológicas, que aproximem os estudantes da realidade do trabalho com IA em ambientes costeiros e oceânicos.

O Risco da Substituição e o Valor do Conhecimento Situado

Embora a automação traga ganhos significativos em escala, velocidade e acurácia, é preciso cautela diante do risco de substituição indiscriminada de competências humanas. O conhecimento empírico dos oceanógrafos — especialmente aquele adquirido em campo, na observação direta de fenômenos marinhos e na interação com comunidades costeiras — continua sendo insubstituível. A sensibilidade para interpretar anomalias, a compreensão dos contextos socioambientais e a capacidade de dialogar com saberes locais ainda não podem ser reproduzidas por máquinas.

Portanto, ao invés de imaginar uma substituição total de humanos por sistemas algorítmicos, o futuro do trabalho na oceanografia deve se orientar pela lógica da complementaridade inteligente: máquinas ampliando a capacidade humana, e não a eliminando.

Desigualdades Tecnológicas e o Acesso à Formação

Um ponto crítico nesse processo de transformação é o acesso desigual às tecnologias e às oportunidades de capacitação. Instituições de ensino e pesquisa com menor infraestrutura tecnológica correm o risco de ficarem à margem da revolução digital oceânica. Isso pode acentuar disparidades regionais e internacionais na produção de conhecimento marinho, além de limitar o papel de cientistas do Sul Global nos fóruns decisórios internacionais relacionados ao oceano.

Assim, é necessário que políticas públicas e agências de fomento priorizem o fortalecimento de capacidades técnico-científicas em regiões historicamente subfinanciadas, garantindo acesso a equipamentos, softwares, bolsas de formação e parcerias estratégicas com centros de excelência em IA aplicada ao oceano.

A Construção de um Oceanógrafo do Século XXI

Em suma, o futuro do trabalho na oceanografia exige mais do que uma atualização curricular — requer uma verdadeira mudança de paradigma. O profissional do século XXI deverá ser simultaneamente cientista, programador, analista crítico e articulador entre diferentes campos do saber. A automação não elimina o papel do oceanógrafo, mas redefine profundamente suas atribuições, suas responsabilidades e suas ferramentas.

Cabe às universidades, aos centros de pesquisa e às instituições reguladoras liderar esse processo de transformação, garantindo que ele ocorra de maneira ética, inclusiva e alinhada às necessidades da sociedade e da sustentabilidade marinha.

CONSIDERAÇÕES FINAIS

A IA não é apenas uma ferramenta a serviço da oceanografia moderna — ela representa uma nova linguagem científica. Sua incorporação crítica e estratégica aos desafios oceânicos pode impulsionar a conservação marinha, a sustentabilidade da economia azul e o avanço do conhecimento em escalas antes inimagináveis. No entanto, como toda tecnologia poderosa, requer discernimento ético, engajamento interdisciplinar e políticas públicas que assegurem sua utilização para o bem comum.

Assim, acredita-se que o futuro da oceanografia será, inevitavelmente, moldado por uma simbiose entre a inteligência humana e a artificial. Cabe à comunidade científica orientar esse processo com responsabilidade, criatividade e compromisso com o oceano e com as sociedades que dele dependem.

REFERÊNCIAS

- ALBAHRA, S. *et al.* Artificial intelligence and machine learning overview in pathology & laboratory medicine: a general review of data preprocessing and basic supervised concepts. **Seminars in Diagnostic Pathology**, v.40,n.2,p.71–8,2023.
- ARDANZA, A. *et al.* Sustainable and flexible industrial human machine interfaces to support adaptable applications in the industry 4.0 paradigm. **International Journal of Production Research**, v.5,n.12,p.4045–4059,2019.
- BACHIMANCH, R. *et al.* Análise de dados com tecnologia de aprendizado profundo. **Cartas de Limnologia e Oceanografia**, v.9,p.324–339,2024.
- GRAY, A. *et al.* The promise and pitfalls of machine learning in ocean remote sensing. **Oceanography**, v.3,n.3,p.52–63,2024.
- HASSAN, M. *et al.* Local interpretable model-agnostic explanation approach for medical imaging analysis: a systematic literature review. **Computers in Biology and Medicine**, v.185,109569,2025.
- KHAN, S. A. *et al.* Integrating industry 4.0 for enhanced sustainability: pathways and prospects. **Sustainable Production and Consumption**, v.54, p.149–189,2025.
- KOOLI, C. *et al.* Chatbots in education and research: a critical examination of ethical implications and solutions. **Sustainability**, v.15,n.7,5614,2023.
- LI, Y. *et al.* NASA NeMO-Net's convolutional neural network: mapping marine habitats with spectrally heterogeneous remote sensing imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v.13,2020.

PERGANTIS, M. *et al.* AI chatbots and cognitive control: enhancing executive functions through chatbot interactions: a systematic review. **Brain Sciences**, v.15,n.1,47,2025.

SADAIAPPAN, B. *et al.* Applications of machine learning in chemical and biological oceanography. **ACS Omega**, v.8,n.18,p.15831–15853,2023.

SCHWING, F. B. Modern technologies and integrated observing systems are “instrumental” to fisheries oceanography: a brief history of ocean data collection. **Fisheries Oceanography**, 2022. <https://doi.org/10.1111/fog.12619>.

ZHAO, Z. *et al.* Applications of deep learning in physical oceanography: a comprehensive review. **Frontiers in Marine Science**, 2024.

INTELIGÊNCIA ARTIFICIAL NA AQUICULTURA: CAMINHOS PARA UMA PRODUÇÃO SUSTENTÁVEL E AUTOMATIZADA

Erick Paiva Barbieri

Universidade Federal Tecnológica do Paraná (UTFPR)

Edison Barbieri

Instituto de Pesca-Governo do Estado de São Paulo

RESUMO

A aquicultura, enquanto setor estratégico para a segurança alimentar global e preservação dos estoques pesqueiros naturais, enfrenta desafios relacionados à sustentabilidade, eficiência e monitoramento ambiental. A Inteligência Artificial (IA), com seu potencial disruptivo, tem se consolidado como ferramenta essencial na transição para sistemas de produção automatizados e baseados em dados. Este artigo explora como algoritmos de aprendizado de máquina, visão computacional e técnicas de processamento de linguagem natural estão sendo empregados para otimizar processos aquícolas, minimizar impactos ecológicos e ampliar a rastreabilidade e o controle sanitário. Além disso, discute-se o papel ético da automação e suas implicações para o trabalho humano nas cadeias produtivas da aquicultura.

Palavras-chave: aquicultura; inteligência artificial; sustentabilidade; aprendizado de máquina; automação.

INTRODUÇÃO

Com o avanço da Indústria 4.0, a integração de tecnologias digitais nas cadeias produtivas tem revolucionado diversos setores da economia global, promovendo a convergência entre automação, análise de dados, inteligência artificial (IA) e conectividade em tempo real. A aquicultura, setor fundamental para a segurança alimentar e o desenvolvimento socioeconômico de regiões costeiras e interiores, não é exceção a esse processo transformador. Atualmente, mais de 50% do pescado consumido mundialmente é proveniente da aquicultura, conforme relatórios recentes da Organização das Nações Unidas para a Alimentação e a Agricultura (FAO, 2022). Essa taxa de crescimento, entretanto, traz consigo importantes desafios relacionados à sustentabilidade ambiental, bem-estar animal, biossegurança e eficiência operacional.

O aumento da demanda global por pescado, aliado à intensificação dos sistemas de cultivo, impõe pressões significativas sobre os recursos naturais, especialmente em relação ao uso de água, insumos zootécnicos e à geração de efluentes. Ao mesmo tempo, cresce a exigência por sistemas de produção que assegurem rastreabilidade, monitoramento contínuo, controle de qualidade e capacidade de resposta rápida a eventos sanitários. Diante desse cenário, a Inteligência Artificial tem se destacado como uma ferramenta promissora para lidar com a complexidade e a dinamicidade dos sistemas aquícolas modernos.

A aplicação de algoritmos de aprendizado de máquina, visão computacional, análise preditiva e processamento de linguagem natural (NLP) permite otimizar a gestão da produção, antecipar riscos, reduzir perdas e promover uma tomada de decisão mais precisa, baseada em grandes volumes de dados (big data). Esses recursos contribuem não apenas para o aumento da produtividade e da rentabilidade, mas também para a mitigação dos impactos ambientais e a promoção de práticas mais éticas e transparentes no cultivo de organismos aquáticos.

Entretanto, a adoção de tecnologias inteligentes na aquicultura não ocorre de forma isenta de obstáculos. Há desafios técnicos relacionados à calibração de sensores, interoperabilidade de sistemas, qualidade dos dados e adaptação a diferentes escalas produtivas (**AMUNDSEN et al., 2021**). Além disso, surgem questões de ordem ética e social, como a substituição da força de

trabalho humana por sistemas automatizados, a exclusão digital de pequenos produtores e o risco de dependência de plataformas tecnológicas proprietárias.

Historicamente, a aquicultura brasileira evoluiu de forma expressiva a partir da década de 1990, consolidando-se nas últimas décadas como uma atividade estratégica para o desenvolvimento rural, a segurança alimentar e a geração de empregos. A tilapicultura, em particular, tornou-se o principal segmento do setor, representando mais de 60% da produção nacional de pescado cultivado. No entanto, o país também possui grande potencial no cultivo de espécies nativas de alto valor ecológico, cultural e comercial, como o tambaqui (*Colossoma macropomum*) e o pirarucu (*Arapaima gigas*), principalmente na Região Norte. De acordo com os dados mais recentes do IBGE (2022), a produção aquícola nacional ultrapassou 800 mil toneladas anuais, com destaque para os estados do Paraná, Rondônia, São Paulo e Maranhão.

Apesar do crescimento contínuo, o setor aquícola brasileiro ainda enfrenta desafios estruturais, como a escassez de dados sistematizados, dificuldades logísticas, carência de assistência técnica e barreiras para a adoção de tecnologias inovadoras, especialmente por pequenos e médios produtores. Nesse cenário, a aplicação de soluções baseadas em IA surge como uma oportunidade concreta para superar gargalos históricos da produção, promovendo maior eficiência no uso de recursos, maior previsibilidade na tomada de decisão e melhor integração entre ciência, tecnologia e práticas produtivas sustentáveis.

As pesquisas brasileira tem desempenhado um papel central na investigação e desenvolvimento de tecnologias voltadas à aquicultura de precisão. Universidades públicas e institutos de pesquisas como o Instituto de Pesca de São Paulo, em parceria com centros de pesquisa como a Embrapa Pesca e Aquicultura, têm gerado conhecimento técnico-científico de ponta, explorando desde o uso de sensores remotos até modelos baseados em inteligência artificial para previsão de crescimento, detecção de anomalias e automação da alimentação. Ao mesmo tempo, começam a surgir startups e iniciativas de inovação aberta voltadas ao setor aquícola, promovendo a transferência de tecnologia e fomentando soluções aplicadas ao contexto brasileiro.

Nesse panorama, políticas públicas bem estruturadas são essenciais para fomentar a adoção dessas inovações, reduzindo desigualdades tecnológicas entre grandes e pequenos produtores. A criação de marcos regulatórios que

estimulem o uso ético e transparente da IA, o financiamento de projetos de base tecnológica por agências como CNPq e FINEP e a articulação de parcerias público-privadas são caminhos estratégicos para impulsionar a transformação digital da aquicultura nacional, garantindo sustentabilidade, competitividade e inclusão produtiva.

O presente capítulo tem como objetivo apresentar um panorama abrangente e crítico das aplicações da IA na aquicultura, abordando suas potencialidades, limitações e implicações para o futuro do setor. Alinha-se às diretrizes temáticas da coletânea "Inteligência Artificial Aplicada: Soluções para um Mundo Automatizado", buscando contribuir com uma reflexão fundamentada sobre os rumos da produção aquícola diante da transformação digital. Serão discutidas as principais inovações tecnológicas em curso, os desafios técnicos enfrentados por pesquisadores e produtores, bem como as implicações sociais, ambientais e econômicas associadas ao uso dessas ferramentas emergentes

APLICAÇÕES DE IA NA AQUICULTURA

A aplicação da Inteligência Artificial na aquicultura tem evoluído rapidamente, impulsionada pela crescente disponibilidade de dados, avanço nos sensores e dispositivos IoT (Internet das Coisas), além da maior capacidade computacional para processamento e análise em tempo real (**GOKULNATH et al., 2024**). A seguir, são discutidas as principais frentes tecnológicas em que a IA tem sido empregada no setor aquícola, com ênfase nas inovações já em curso e nas perspectivas de desenvolvimento.

Monitoramento Ambiental Automatizado

Sistemas baseados em IA têm sido empregados para o monitoramento contínuo de variáveis ambientais críticas, como temperatura, oxigênio dissolvido, pH, salinidade, turbidez e concentração de amônia. Através de sensores interligados a redes neurais artificiais, é possível detectar padrões de variação que precedem eventos de estresse ou mortalidade em massa dos organismos cultivados, permitindo intervenções precoces e mais eficientes.

O uso de algoritmos preditivos, como Random Forest, Support Vector Machines (SVM) e redes neurais profundas (Deep Learning), tem permitido o

desenvolvimento de modelos que correlacionam variáveis ambientais a respostas fisiológicas dos organismos, antecipando surtos de doenças, colapsos de oxigênio e episódios de eutrofização (BAO et al., 2020). Em ambientes de cultivo intensivo, essas ferramentas têm contribuído para a redução de perdas econômicas e para o aprimoramento do manejo ambiental.

Alimentação de Precisão com Visão Computacional

A alimentação representa entre 50% e 70% dos custos operacionais em sistemas aquícolas, sendo um dos pontos críticos para a viabilidade econômica da atividade. Nesse contexto, a utilização de sistemas de alimentação inteligente, baseados em visão computacional e aprendizado de máquina, tem revolucionado o manejo nutricional.

Por meio de câmeras subaquáticas e softwares capazes de interpretar imagens em tempo real, os sistemas identificam o comportamento alimentar dos peixes e ajustam automaticamente a oferta de ração. Tais sistemas reduzem significativamente o desperdício de alimento, melhoram a conversão alimentar e diminuem a carga orgânica nos tanques e viveiros, contribuindo para a sustentabilidade ambiental da produção.

Algoritmos de IA têm sido utilizados para ajustar a formulação nutricional de dietas com base em dados fisiológicos, desempenho zootécnico e variações climáticas, promovendo uma abordagem dinâmica e personalizada da nutrição.

Diagnóstico de Doenças e Detecção de Anomalias

A identificação precoce de doenças é fundamental para garantir o bem-estar animal e reduzir o uso de antimicrobianos na aquicultura. Sistemas baseados em IA têm sido empregados para analisar sinais comportamentais, imagens, vídeos e dados biométricos dos organismos, detectando alterações que indicam infecções, parasitismos ou deficiências nutricionais.

Modelos de aprendizado supervisionado treinados com grandes bases de dados clínicos permitem a detecção automática de patologias específicas, com índices de acurácia superiores aos obtidos por métodos convencionais. Algumas plataformas integram IA com sensores de biomassa, permitindo a

correlação entre oscilações populacionais e surtos patológicos, o que representa um avanço na gestão sanitária preventiva (**BARBOSA et al., 2022**).

Rastreamento e Gestão Inteligente da Produção

A rastreabilidade é uma exigência crescente no comércio internacional de pescado, sendo fundamental para garantir a segurança alimentar e a conformidade com padrões sanitários e ambientais. A IA tem desempenhado um papel estratégico na automação da coleta, organização e análise de dados ao longo de toda a cadeia produtiva.

Soluções de blockchain integradas a sistemas de IA permitem rastrear, em tempo real, informações sobre origem, manejo, transporte, qualidade da água, alimentação e sanidade dos animais. Essa integração facilita auditorias, gera maior transparência para o consumidor final e fortalece a reputação dos produtores em mercados exigentes.

Sistemas baseados em IA podem sugerir ajustes no manejo com base em previsões de crescimento, demanda de mercado e condições climáticas, promovendo uma gestão adaptativa e mais eficiente.

Seleção Genética Assistida por IA

A aplicação de técnicas de IA em programas de melhoramento genético tem sido promissora para acelerar a identificação de indivíduos com características zootécnicas superiores, como crescimento rápido, resistência a doenças, conversão alimentar eficiente e tolerância a variabilidade ambiental.

Modelos baseados em algoritmos genéticos, aprendizado profundo e análise multivariada têm sido utilizados para processar grandes volumes de dados genômicos e fenotípicos, identificando padrões e loci de interesse com alta acurácia. Essa abordagem tem contribuído para a criação de linhagens mais resilientes, adaptadas a diferentes biomas e condições produtivas.

ÉTICA, SUSTENTABILIDADE E O FUTURO DO TRABALHO NA AQUICULTURA AUTOMATIZADA

Apesar do grande potencial da Inteligência Artificial na otimização dos sistemas aquícolas, sua aplicação enfrenta diversos desafios que vão além do

domínio tecnológico. Tais limitações envolvem aspectos técnicos, estruturais, econômicos, regulatórios e éticos, que precisam ser considerados para que a transição digital do setor ocorra de forma sustentável, inclusiva e responsável.

Limitações Técnicas e Infraestruturais

A implementação de sistemas baseados em IA demanda infraestrutura adequada, como conectividade de alta qualidade, disponibilidade de energia elétrica estável, sensores de precisão e servidores capazes de processar grandes volumes de dados em tempo real (PRASANNA KUMAR *et al.*, 2025). Em muitas regiões produtoras, especialmente em países em desenvolvimento, essas condições ainda são precárias ou inexistentes, o que compromete a efetividade e a escalabilidade das soluções propostas.

A calibração de algoritmos requer conjuntos robustos e bem estruturados de dados históricos, os quais nem sempre estão disponíveis. A ausência de padronização nos registros, a heterogeneidade dos sistemas produtivos e a variabilidade genética das espécies cultivadas são fatores que limitam a generalização dos modelos preditivos e dificultam sua aplicação ampla.

Barreiras Econômicas e Desigualdade Tecnológica

O custo de aquisição e manutenção de tecnologias baseadas em IA, bem como o treinamento de pessoal qualificado para operá-las, representa um obstáculo significativo, principalmente para pequenos e médios produtores. A centralização das inovações em grandes conglomerados aquícolas pode acentuar ainda mais as desigualdades tecnológicas, excluindo agricultores familiares e cooperativas da transição digital.

Programas de incentivo público, políticas de financiamento acessível e parcerias público-privadas são essenciais para democratizar o acesso à IA e promover sua adoção em diferentes escalas de produção, garantindo que os benefícios da automação não fiquem restritos a grandes empreendimentos.

Questões Éticas e Responsabilidade Algorítmica

O uso de sistemas automatizados levanta importantes questões éticas, especialmente no que se refere à transparência dos algoritmos, à proteção de

dados sensíveis dos produtores e à responsabilidade por decisões tomadas por sistemas autônomos. A opacidade de certos modelos de aprendizado profundo (conhecidos como “caixas-pretas”) dificulta a compreensão e a auditoria dos critérios utilizados para inferências, o que pode comprometer a confiança dos usuários.

Faz-se necessário discutir os impactos sobre o emprego e as condições de trabalho. Embora a automação possa liberar os trabalhadores de tarefas repetitivas e extenuantes, também há riscos de substituição de mão de obra sem o devido planejamento de requalificação profissional, aprofundando a exclusão social e econômica de determinados grupos.

Sustentabilidade e Conformidade Regulatória

A adoção de tecnologias digitais deve estar alinhada aos princípios da sustentabilidade ambiental, econômica e social. Ainda são incipientes os marcos regulatórios que orientam o uso ético e ambientalmente responsável da IA na aquicultura. A ausência de diretrizes claras dificulta o licenciamento de novas tecnologias e pode gerar insegurança jurídica, tanto para empreendedores quanto para órgãos fiscalizadores.

É fundamental que legislações e normativas técnicas acompanhem o ritmo da inovação, estabelecendo parâmetros de segurança, interoperabilidade, compartilhamento de dados e avaliação de riscos, assegurando que as soluções digitais estejam em conformidade com as boas práticas ambientais e com os direitos humanos.

LINHAS FUTURAS DE PESQUISA EM IA PARA AQUICULTURA: RUMO À INOVAÇÃO SUSTENTÁVEL

A Inteligência Artificial (IA) está transformando a aquicultura, tornando-a mais eficiente, sustentável e resiliente (DONG *et al.*, 2020). Este campo emergente oferece vastas oportunidades de pesquisa para enfrentar desafios globais como mudanças climáticas, demanda crescente por alimentos e conservação da biodiversidade aquática.

IA e Bem-Estar Animal Avançado

Objetivo: Desenvolver sistemas inteligentes de monitoramento, utilizando redes neurais profundas (Deep Learning), para identificar e avaliar indicadores de bem-estar animal em organismos aquáticos, como peixes e crustáceos. O sistema buscará detectar sinais precoces de estresse, alterações comportamentais e comprometimentos da saúde mental dos animais, por meio da análise integrada de dados comportamentais, fisiológicos e ambientais. Essa abordagem se alinha às exigências contemporâneas de sustentabilidade e bioética, conforme discutido por Somdutt (2024).

Aplicações:

Análise de padrões comportamentais: Utilização de visão computacional subaquática e algoritmos de aprendizado profundo para reconhecer padrões de nado, alimentação, interações e tempo de inatividade. Esses dados serão analisados em tempo real para inferir o estado emocional e fisiológico dos animais, detectando desvios comportamentais que indiquem sofrimento ou disputas hierárquicas disfuncionais.

Detecção precoce de doenças e estressores ambientais: Modelos preditivos poderão correlacionar comportamentos atípicos com infecções ou exposição a contaminantes. Isso permitirá intervenções antes da manifestação visual dos sintomas, reduzindo a mortalidade, além de integrar dados ambientais (pH, oxigênio dissolvido) para entender os impactos nas respostas comportamentais.

Integração com biossensores e IoT: Sensores não invasivos conectados à Internet das Coisas (IoT) poderão coletar dados fisiológicos contínuos, como cortisol dissolvido e taxa respiratória. Esses dados serão integrados aos modelos de Deep Learning para fornecer avaliações holísticas e em tempo real sobre o bem-estar animal.

Desafios e Considerações Éticas:

Coleta de dados em larga escala: A diversidade entre espécies de peixes e crustáceos exige bancos de dados padronizados e representativos, com diferentes faixas etárias e contextos ambientais.

Antropomorfização: A interpretação de comportamentos animais com base em categorias humanas exige cautela para evitar erros de julgamento e decisões inadequadas.

Aquicultura de Precisão com Edge Computing e 5G

Objetivo: Modernizar a aquicultura por meio da implementação de tecnologias de Edge Computing e redes 5G, possibilitando a coleta, processamento e análise de dados diretamente nos dispositivos de campo. Isso reduz a latência, diminui a dependência de servidores em nuvem e oferece respostas imediatas a variáveis críticas da produção, como alimentação, saúde animal e qualidade da água. A descentralização do processamento de dados melhora a eficiência operacional e favorece a sustentabilidade, especialmente em regiões com conectividade limitada.

Aplicações:

Alimentação automatizada e responsiva: Sistemas de alimentação inteligente, baseados em algoritmos de aprendizado federado, ajustam as doses de ração em tempo real, com base em variáveis como temperatura da água, comportamento alimentar dos peixes e parâmetros fisiológicos monitorados localmente. Essa abordagem reduz desperdícios, melhora o índice de conversão alimentar (ICA) e diminui a carga orgânica no sistema, promovendo um manejo mais eficiente e ambientalmente sustentável.

Diagnóstico descentralizado e acessível: Ferramentas diagnósticas embarcadas em dispositivos móveis permitirão que pequenos e médios produtores identifiquem precocemente patologias como infecções e parasitoses. Aplicativos com inteligência artificial permitirão triagens clínicas com base em imagens, vídeos ou dados do ambiente, sem a necessidade de conexão constante com servidores centrais, democratizando o acesso a tecnologias avançadas.

Redes de sensores autossustentáveis e conectividade robusta: Sensores ambientais de baixo consumo energético, alimentados por painéis solares e conectados por redes LoRaWAN, monitorarão variáveis críticas como oxigênio dissolvido e pH, transmitindo dados para os pontos de borda. A integração

com infraestrutura 5G expandirá o número de sensores e permitirá a tomada de decisões distribuídas em tempo real.

Desafios e Limitações Técnicas:

Conectividade em regiões precárias: A implementação dessas tecnologias em áreas com infraestrutura limitada exige soluções adaptadas, como antenas de longo alcance, redes mesh locais e dispositivos que funcionem offline, com sincronização periódica.

Redução de custos e acessibilidade: A adoção em larga escala depende da redução dos custos de hardware, software e manutenção, além de parcerias público-privadas e modelos cooperativos para tornar as tecnologias viáveis para pequenos aquicultores.

Capacitação técnica e inclusão digital: Programas de capacitação serão essenciais para garantir que os produtores possam interpretar dados e interagir com sistemas automatizados.

Modelos Preditivos para Mudanças Climáticas

Objetivo: Desenvolver e aplicar sistemas baseados em inteligência artificial generativa, com foco em redes adversariais generativas (GANs), para prever eventos climáticos extremos e formular estratégias de adaptação na aquicultura frente às mudanças ambientais globais. Esses sistemas visam antecipar alterações críticas nos padrões ambientais, como temperatura, salinidade, acidificação e hipóxia, além de propor cenários resilientes e medidas adaptativas, promovendo a sustentabilidade da produção aquícola, especialmente em regiões vulneráveis.

Aplicações:

Sistemas de alerta precoce para eventos extremos: A modelagem preditiva com GANs permite detectar condições ambientais adversas iminentes, como ondas de calor marinhas, tempestades e episódios de acidificação. Integrados a plataformas de monitoramento remoto e sensores, esses sistemas emitem alertas em tempo real, permitindo que os produtores adotem medidas preventivas, como ajuste no manejo alimentar ou acionamento de sistemas de aeração de emergência.

Simulação de cenários de cultivo resilientes: Os modelos generativos criam dados sintéticos para simular possíveis trajetórias climáticas futuras, testando diferentes estratégias de cultivo adaptativo, como variações no tempo de despesca, seleção de espécies tolerantes a estresses e ajustes no sistema de cultivo (ex. RAS). A simulação de condições futuras é uma ferramenta valiosa para o planejamento de médio e longo prazo na aquicultura.

Geointeligência e planejamento territorial: A combinação de dados geoespaciais, sensores ambientais e IA generativa permite identificar áreas propícias à instalação de empreendimentos aquícolas resilientes. A análise de múltiplas camadas de dados, como batimetria, temperatura superficial do mar e histórico de eventos climáticos, promove uma ocupação territorial mais estratégica e segura.

Desafios e Considerações Estratégicas:

Integração de dados multiescala: A implementação eficaz exige integrar dados de escalas espaciais e temporais distintas, de microclimas locais a projeções globais de mudanças climáticas, o que demanda soluções computacionais robustas.

Adaptação a zonas tropicais e subtropicais: A regionalização dos modelos preditivos é necessária, incluindo dados empíricos locais, para melhorar a acurácia em ecossistemas tropicais e subtropicais.

Ética e confiabilidade: A geração de dados sintéticos com GANs deve ser validada por especialistas, garantindo que os cenários propostos estejam alinhados com a realidade ecológica e socioeconômica local.

IA PARA ESPÉCIES NÃO CONVENCIONAIS E CONSERVAÇÃO

O uso de redes neurais convolucionais (CNNs), uma arquitetura de aprendizado profundo eficaz para processamento de imagens e reconhecimento de padrões visuais, visa otimizar o estudo, monitoramento e conservação de espécies nativas ameaçadas. Essa abordagem inovadora busca superar limitações tradicionais de amostragem e monitoramento, fornecendo ferramentas precisas para apoiar políticas de manejo e proteção ambiental, essenciais para a preservação da biodiversidade e a exploração sustentável.

Aplicações:

Identificação automatizada de indivíduos: As CNNs permitem o reconhecimento não invasivo de indivíduos com base em características morfológicas, como padrões de coloração ou marcas corporais. Integradas a câmeras fixas ou móveis (incluindo drones ou veículos subaquáticos), essas tecnologias possibilitam o monitoramento contínuo de populações em habitats naturais, evitando o estresse associado à captura de animais.

Acompanhamento de ciclos reprodutivos e desova: A integração de CNNs com drones ou câmeras em pontos críticos permite monitorar com alta resolução temporal e espacial os ciclos reprodutivos de espécies como os bagres amazônicos. Essas ferramentas ajudam a mapear habitats de reprodução, identificar alterações comportamentais e avaliar o sucesso de estratégias de conservação.

Combate à pesca ilegal: A instalação de câmeras inteligentes equipadas com CNNs em áreas protegidas permite detectar atividades suspeitas, como embarcações não autorizadas ou redes de pesca ilegais. Esses sistemas podem gerar alertas em tempo real, aumentando a eficiência da fiscalização e protegendo as populações vulneráveis.

Desafios e Limitações:

Escassez de dados para treinamento: A falta de bancos de dados de imagens rotuladas, especialmente em regiões tropicais, limita a eficácia dos modelos. Estratégias como data augmentation, transferência de aprendizado e a criação de bancos colaborativos de imagens podem ajudar a superar essa barreira.

Conciliação entre exploração econômica e conservação: Embora as CNNs permitam uma exploração mais eficiente de espécies, é fundamental garantir que essa exploração não comprometa os ecossistemas naturais. Isso exige regulamentação rigorosa, avaliação ambiental contínua e o engajamento de comunidades locais e gestores ambientais.

Blockchain e IA para Rastreabilidade Ética

Objetivo:

Assegurar a sustentabilidade ambiental e a rastreabilidade operacional das atividades aquícolas por meio da adoção integrada de tecnologias emergentes, como a Inteligência Artificial (IA) e o blockchain. A proposta visa não apenas otimizar a eficiência da produção, mas também garantir transparência, confiabilidade e conformidade com padrões internacionais de boas práticas ambientais e sociais (ZHOU *et al.*, 2017).

Aplicações Estratégicas:

Certificações automatizadas por meio de contratos inteligentes: O uso de *smart contracts*, baseados em plataformas blockchain, possibilita o registro imutável e automatizado de práticas sustentáveis adotadas nas unidades de cultivo aquícola. Esses contratos podem ser programados para emitir certificações ambientais, sociais e de qualidade apenas quando determinados critérios, previamente acordados entre as partes interessadas, forem atendidos. Tal abordagem mitiga a possibilidade de fraudes e aumenta a confiança entre produtores, consumidores e órgãos reguladores.

Análise de conformidade com Processamento de Linguagem Natural (PLN): A aplicação de algoritmos de PLN permite a leitura e interpretação automatizada de relatórios de auditoria, registros operacionais e documentos técnicos. Esses sistemas podem identificar inconsistências, não conformidades e padrões linguísticos que indiquem potenciais violações às normas de sustentabilidade. Ao reduzir a necessidade de intervenção humana contínua, essa aplicação torna o processo de auditoria mais ágil, econômico e escalável, inclusive em áreas remotas ou de difícil acesso.

Tokenização de créditos de carbono no mercado voluntário: A tokenização de créditos de carbono representa uma inovação no financiamento ambiental. Por meio de blockchain, é possível transformar os créditos gerados por práticas aquícolas de baixa emissão em ativos digitais negociáveis. Isso promove a inserção das fazendas aquícolas em mercados de carbono voluntários, fomentando incentivos econômicos concretos à adoção de tecnologias

limpas, como sistemas de recirculação de água, uso de energia renovável e manejo sustentável de resíduos.

Desafios para a Implementação:

Interoperabilidade entre sistemas de informação: Um dos maiores entraves para a aplicação dessas tecnologias em escala global reside na dificuldade de integração entre diferentes plataformas digitais, regulamentos locais e protocolos técnicos utilizados em distintas jurisdições. É imperativo o desenvolvimento de padrões internacionais de interoperabilidade que permitam a comunicação fluida entre sistemas heterogêneos, respeitando a soberania dos dados e as legislações nacionais.

Confiança e aceitação do consumidor final: Embora as tecnologias digitais possam aumentar a transparência e a rastreabilidade, sua efetividade depende da aceitação por parte do consumidor. Muitos consumidores ainda demonstram ceticismo quanto à confiabilidade das certificações digitais e à complexidade dos sistemas baseados em blockchain. Estratégias de comunicação clara, campanhas de educação ambiental e rotulagem inteligente podem ajudar a construir essa confiança, tornando o consumidor um aliado na promoção de práticas aquícolas sustentáveis.

Robótica Autônoma em Aquicultura Offshore

Objetivo:

O desenvolvimento de robôs subaquáticos autônomos com sistemas de Inteligência Artificial distribuída, baseados na lógica de enxames (swarm intelligence), representa uma inovação crucial para a aquicultura offshore. O objetivo é criar plataformas robóticas capazes de operar de forma cooperativa em ambientes marinhos complexos, realizando tarefas críticas sem intervenção humana direta. Essas soluções buscam aumentar a eficiência, reduzir custos operacionais, minimizar riscos à segurança humana e promover a sustentabilidade ambiental.

Aplicações Estratégicas:

Manutenção de infraestruturas submersas: Robôs autônomos equipados com sensores multiespectrais, câmeras subaquáticas e braços manipuladores

podem realizar tarefas de limpeza, detecção de falhas e reparos emergenciais. Com coordenação baseada em IA swarm, múltiplos robôs podem atuar simultaneamente em áreas extensas, otimizando o tempo e os recursos.

Colheita seletiva e manejo de biomassa: Sensores fisiológicos e câmeras permitem que robôs identifiquem peixes e moluscos com base em seu tamanho e comportamento, realizando colheitas seletivas com precisão e baixo impacto ambiental. Essa abordagem contribui para o bem-estar animal e para certificações de responsabilidade socioambiental.

Inspeção de estruturas em águas profundas: Robôs autônomos capazes de suportar altas pressões e salinidade, equipados com sensores como ultrassom e LIDAR subaquático, podem mapear em tempo real a integridade das instalações submersas. Eles detectam fissuras, deformações e bioincrustações, sendo essenciais para áreas afastadas da costa, onde o acesso humano é limitado ou perigoso.

Desafios Tecnológicos e Regulatórios:

Materiais avançados e resistentes: A corrosão e a degradação por salinidade e pressão exigem o uso de materiais como titânio, aço inoxidável e revestimentos antifouling. Os sistemas eletrônicos devem ser altamente vedados para garantir resistência e durabilidade.

Normas e regulamentações: O uso de robôs em ambientes marinhos compartilhados necessita de regulamentações específicas para garantir segurança, proteção da biodiversidade e interoperabilidade com sistemas humanos. A governança deve envolver diferentes setores, assegurando transparência e controle social.

IA Inclusiva para Pequenos Produtores

Objetivo: Este eixo temático busca democratizar o acesso às tecnologias de Inteligência Artificial (IA), com foco em soluções de baixo custo, fácil usabilidade e sensíveis à cultura local, voltadas para pequenos produtores, pescadores artesanais e comunidades tradicionais. O objetivo é garantir que os avanços da IA melhorem a produção aquícola, promovendo equidade, soberania

tecnológica e o protagonismo de populações historicamente marginalizadas no processo de transformação digital.

Aplicações Estratégicas:

Chatbots locais para extensão rural e assistência técnica: Assistentes virtuais, acessíveis por smartphones ou rádios comunitários e treinados em línguas nativas, podem transformar a extensão rural em comunidades remotas. Esses sistemas de IA forneceriam informações sobre práticas de manejo, previsão do tempo, preços de mercado, nutrição e normativas ambientais, promovendo a difusão de conhecimento técnico de forma contínua, personalizada e intercultural. Também podem ser usados em campanhas educativas sobre saúde pública e direitos territoriais.

Diagnóstico por imagem com aprendizado de máquina: Aplicativos móveis, operando offline e equipados com aprendizado de máquina, permitiriam que pequenos produtores identificassem doenças em peixes, moluscos e crustáceos a partir de fotografias tiradas no campo. Essa tecnologia reduziria a dependência de técnicos especializados e melhoraria a resposta a surtos sanitários, minimizando perdas econômicas.

Plataformas colaborativas e redes comunitárias de dados: Plataformas digitais de IA com design participativo e código aberto podem criar redes de dados entre comunidades produtoras, promovendo a gestão cooperativa dos recursos naturais e a construção de indicadores de sustentabilidade. Essas plataformas integrariam dados meteorológicos, de qualidade da água e de produção, auxiliando na tomada de decisões coletivas e fortalecendo a governança local.

Desafios Técnicos, Culturais e Éticos:

Superar barreiras culturais, linguísticas e educacionais: A adoção de IA deve ser co-desenhada com as comunidades, respeitando seus saberes tradicionais e línguas. Soluções como energia solar e servidores locais são necessárias para superar limitações de infraestrutura. A formação de multiplicadores comunitários é essencial para garantir a adoção eficaz das tecnologias.

Proteger a privacidade e soberania dos dados: É fundamental garantir que os dados das comunidades permaneçam sob seu controle, respeitando os

princípios de Consentimento Livre, Prévio e Informado (CLPI) e os direitos previstos pela OIT. A governança dos algoritmos deve ser transparente e baseada em marcos éticos sólidos

CONSIDERAÇÕES FINAIS

A integração da Inteligência Artificial (IA) à aquicultura marca um avanço significativo na transformação digital do setor, com impactos relevantes sobre a produtividade, a sustentabilidade ambiental e o bem-estar animal. Tecnologias inteligentes permitem o monitoramento em tempo real, o controle automatizado de variáveis críticas e decisões orientadas por dados, otimizando o uso de recursos e reduzindo desperdícios.

No entanto, a ampla adoção da IA ainda enfrenta desafios técnicos, econômicos, regulatórios e sociais. A viabilidade de soluções escaláveis e acessíveis requer investimentos em infraestrutura digital, políticas públicas inclusivas e capacitação de produtores e técnicos. Além disso, questões éticas e legais, como o uso responsável de dados, a transparência dos algoritmos e a equidade no acesso às inovações, precisam ser consideradas.

O futuro da aquicultura inteligente dependerá de parcerias interdisciplinares entre especialistas em computação, engenharia, biologia aquática e formulação de políticas. Tecnologias emergentes, como computação em borda, Internet das Coisas aquática e blockchain, prometem impulsionar sistemas adaptativos e resilientes.

IA representa uma revolução silenciosa, porém estruturante, na aquicultura. Seu sucesso exigirá rigor científico, ética e visão estratégica para garantir que seus benefícios sejam amplamente distribuídos e contribuam para a segurança alimentar global.

REFERÊNCIAS

AMUNDSEN *et al.* Autonomous ROV inspections of aquaculture net pens using DVL. *IEEE Journal of Oceanic Engineering*, v.47, p.1–19, 2021. DOI: 10.1109/JOE.2021.3105285.

BAO et al. Integrated navigation for autonomous underwater vehicles in aquaculture: A review. *Information Processing in Agriculture*, v.7, p.139–151, 2020. DOI: 10.1016/j.inpa.2019.04.003.

BARBOSA et al. High-resolution drone images show that the distribution of mussels depends on microhabitat features of intertidal rocky shores. *Remote Sensing*, v.14, 5441, 2022. DOI: 10.3390/rs14215441.

DONG et al. Application of computer vision in fish intelligent feeding system—A review. *Aquaculture Research*, 2020. DOI: 10.1111/are.14907.

FAO – FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. *The State of World Fisheries and Aquaculture 2022*. Rome: FAO, 2022. Disponível em: <https://www.fao.org/documents/card/en/c/cc0461en>. Acesso em: 24 abr. 2025.

GOKULNATH et al. *Precision Aquaculture: Empowering Fish Farming with AI and IoT*. In: Precision Aquaculture: Empowering Fish Farming with AI and IoT. Elite Publication House, 2024.

PRASANNA et al. Fish disease detection in aquaculture using Pseudo Hamiltonian Neural Network optimized with Philippine Eagle Optimization Algorithm. *Knowledge-Based Systems*, 2025. Art. 113374. DOI: 10.1016/j.knosys.2025.113374.

SOMDUJT, T. AI-powered agriculture chatbots for farmers. In: *Agricultural Extension: Worldwide Innovations*. Elite Publication House, 2024.

ZHOU et al. Intelligent feeding control methods in aquaculture with an emphasis on fish: a review. *Reviews in Aquaculture*, 2017. DOI: 10.1111/raq.12218.

RECONHECIMENTO FACIAL E RACISMO ALGORÍTMICO: OS DESAFIOS NA SEGURANÇA PÚBLICA BRASILEIRA

Roberto Cezar Marcelino Neris
Universidade Tecnológica Federal do Paraná (UTFPR)

RESUMO

O artigo examina os impactos éticos, sociais e técnicos decorrentes do uso de sistemas de reconhecimento facial na segurança pública brasileira, com foco no fenômeno do racismo algorítmico. Embora promovidos como soluções eficazes para o combate ao crime, esses sistemas frequentemente reproduzem desigualdades estruturais devido a vieses presentes nos dados de treinamento, afetando desproporcionalmente grupos racializados. A pesquisa discute os riscos a neutralidade da tecnologia de reconhecimento facial. Rejeitar falsas neutralidades das tecnologias, bases de dados e representações é um esforço frequente contra as opressões e limitações. Além disso, o estudo propõe soluções como auditorias algorítmicas e maior representatividade das bases de dados, bem como diretrizes normativas para garantir uma aplicação responsável e ética da tecnologia. Ao integrar inovação tecnológica e princípios de justiça social, este trabalho visa contribuir para a implementação de sistemas mais confiáveis e justos na segurança pública.

Palavras-chave: reconhecimento facial; algoritmos; inteligência artificial; segurança pública; racismo algorítmico.

INTRODUÇÃO

O reconhecimento facial é uma tecnologia automatizada que identifica indivíduos com base em características faciais, sendo amplamente empregada em setores como vigilância, segurança e identificação criminal. Seu uso tem crescido nos últimos anos, impulsionando também o debate sobre os impactos sociais e éticos dessas ferramentas.

Pesquisas indicam que esses sistemas podem apresentar resultados imprecisos, sobretudo em relação a pessoas negras — fenômeno conhecido como racismo algorítmico. Isso decorre da reprodução de desigualdades históricas por meio de bases de dados enviesadas. No Brasil, onde desigualdades socioeconômicas e raciais são profundas, tais distorções assumem particular gravidade.

A crença na neutralidade da Inteligência Artificial pode obscurecer os efeitos discriminatórios dessas tecnologias. Como aponta Benjamin (2019), algoritmos tidos como neutros podem reforçar estruturas de opressão se aplicados sem uma análise crítica. Diante disso, torna-se necessário abordar tanto os aspectos técnicos quanto os impactos sociais do reconhecimento facial.

Este artigo busca responder à seguinte questão: como reduzir os vieses nos algoritmos usados em sistemas de reconhecimento facial, especialmente no contexto da segurança pública? A análise considera os fatores técnicos que contribuem para o racismo algorítmico, bem como os impactos sociais resultantes de seu uso.

A escolha do tema se justifica pela crescente adoção dessa tecnologia em ambientes públicos no Brasil e pelos riscos que ela representa à equidade e aos direitos fundamentais. Parte-se da hipótese de que a integração de medidas técnicas, como auditorias e diversificação de bases de dados, associadas a políticas públicas, pode mitigar os efeitos discriminatórios.

A metodologia combina revisão crítica da literatura e análise de casos práticos, enfocando os desafios éticos, técnicos e legais. O estudo também observa experiências nacionais e internacionais de regulação e monitoramento dessas tecnologias.

Assim, este artigo busca contribuir para o debate sobre o uso do reconhecimento facial na segurança pública brasileira, oferecendo uma abordagem

crítica e fundamentada sobre seus riscos e possíveis caminhos para uma aplicação mais justa.

EVOLUÇÃO DO RECONHECIMENTO FACIAL E SEUS DESAFIOS NA SEGURANÇA PÚBLICA

A tecnologia de reconhecimento facial remonta à década de 1960, quando pioneiros como Woody Bledsoe, Helen Chan Wolf e Charles Bisson iniciaram pesquisas sobre identificação biométrica. Inicialmente, os sistemas enfrentavam desafios devido às limitações tecnológicas da época, como a baixa capacidade de processamento e a falta de padrões para a captura de imagens (Tuteja, *et al.*, 2014).

Na década de 1970, Goldstein, Harmon e Lesk propuseram 21 marcadores subjetivos para o reconhecimento facial, mas a abordagem encontrou problemas devido à variabilidade nas condições de captura das imagens (Sinha; Poggio, 1996 *apud* Tuteja, *et al.*, 2014). Posteriormente, Kohonen desenvolveu um sistema neural que realizava cálculos de reconhecimento facial em imagens padronizadas, utilizando autovetores conhecidos como "*autofaces*" (Sinha, 2002).

Em 1988, Kirby e Sirovich introduziram o uso do método PCA (representação polinomial padrão) no reconhecimento facial, demonstrando como imagens ajustadas adequadamente eram fundamentais para a precisão do processo (Tuteja *et al.*, 2014). No início da década de 1990, Turk e Pentland aprimoraram essa abordagem ao desenvolverem os "*eigenfaces*", um marco na evolução do reconhecimento facial, permitindo a identificação de rostos em fotografias. O método recebeu ampla atenção após um teste no Super Bowl de 2001, evento que gerou grande interesse da mídia e levantou debates sobre sua aplicação na segurança (Tuteja *et al.*, 2014).

Atualmente, o reconhecimento facial está presente em setores, como segurança pública e autenticação em dispositivos móveis. O avanço da tecnologia tem sido impulsionado pelo uso de técnicas modernas, como Redes Neurais Convolucionais (CNNs), *DeepFace* e *DeepID*, que empregam aprendizado profundo (*Deep Learning*) para a identificação e verificação de imagens faciais (Vishwanatha *et al.*, 2023).

Desde a introdução dos métodos de aprendizado profundo, como o *DeepFace* e o *DeepID*, em 2014, a tecnologia sofreu uma transformação

significativa. O *DeepFace*, introduzido por Schroff *et al.* (2015), utiliza aprendizado de representações unificadas para identificação e agrupamento facial, alcançando altos níveis de precisão. O *DeepID*, desenvolvido por Parkhi *et al.* (2015), explora características discriminativas a partir de múltiplas camadas de CNNs, aprimorando a precisão do reconhecimento facial.

Entretanto, com o avanço da tecnologia e a crescente disponibilidade de dados, surgiram preocupações sobre privacidade e consentimento. Raji e Fried (2020) destacam que modelos baseados em *Deep Learning* estão cada vez mais dispensando a solicitação de consentimento para coleta de imagens faciais. As autoras também ressaltam a crescente complexidade do controle sobre esses dados, especialmente em um contexto de coleta em larga escala.

Com a complexidade crescente da gestão de dados, sistemas de reconhecimento facial têm se consolidado como ferramentas para identificação e verificação de identidade por meio da análise de padrões faciais (Phillips *et al.*, 2018). Algoritmos avançados convertem imagens faciais em representações matemáticas, comparando-as com imagens armazenadas em bancos de dados (Li; Jain, 2011). Segundo esses autores, o processo ocorre em duas etapas principais: extração e seleção de características, seguidas pela classificação dos dados (Petrescu, 2019).

Avanços recentes, como modelos tridimensionais e câmeras térmicas, têm melhorado a precisão dos sistemas (Galterio; Shavit; Hayajneh, 2018; Petrescu, 2019). No campo da vigilância, o reconhecimento facial se destaca como uma tecnologia emergente e promissora (Oliveira *et al.*, 2022). Desde os primeiros estudos na década de 1960, a evolução da tecnologia tem atraído interesse para sua implementação no processamento automatizado de imagens digitais (Raviv, 2020).

Na segurança pública, o reconhecimento facial tem como objetivo aumentar a eficiência das operações policiais, permitindo a identificação de criminosos em tempo real (Araújo; Cardoso; Paula, 2021). No entanto, desafios persistem, incluindo preocupações éticas e legais relacionadas à privacidade e aos vieses raciais. Almeida (2022) observa que taxas de erro são maiores para pessoas negras, levantando questionamentos sobre a precisão e o potencial discriminatório do uso policial dessa ferramenta.

O reconhecimento facial para fins policiais já acumula mais de duas décadas de história. Sua popularização foi impulsionada pelo barateamento das tecnologias, pela expansão das bases de dados biométricos e pela falta de regulamentação adequada (Silva, 2022). O primeiro uso documentado ocorreu nos Estados Unidos, durante o *Super Bowl* de 2001, quando câmeras equipadas com reconhecimento facial identificaram suspeitos em meio à multidão (Li; Jain, 2011). Desde então, a tecnologia tem sido amplamente empregada em estratégias de policiamento preditivo e vigilância em espaços públicos.

Com a adoção crescente da Inteligência Artificial, inúmeras empresas veem enfrentando desafios técnicos e dilemas éticos, frequentemente desencadeando controvérsias sobre o impacto dessas tecnologias (Wen; Holweg, 2024). Embora eficaz em setores como finanças, saúde e educação, sua aplicação na segurança pública exige um debate cuidadoso. Oliveira *et al.* (2022), alertam para a necessidade de soluções robustas que abordem implicações éticas, como discriminação e vigilância massiva. Silva (2022) enfatiza a importância de enfrentar desafios éticos, técnicos e legais para garantir uma aplicação justa dessas ferramentas na segurança pública.

VIESES ALGORÍTMICOS E AS LIMITAÇÕES DA INTELIGÊNCIA ARTIFICIAL

Com a crescente presença da Inteligência Artificial (IA) na vida cotidiana, aumentam as preocupações sobre sua capacidade de reproduzir discriminações e desigualdades. O chamado racismo algorítmico (Silva, 2022) decorre da interação entre sistemas automatizados e o preconceito estrutural da sociedade. Embora percebidos como neutros, esses sistemas refletem dados enviesados e decisões de seus desenvolvedores, o que contribui para a reprodução de injustiças históricas.

Essa problemática, torna-se ainda mais evidente nos algoritmos de reconhecimento facial que, conforme demonstrado por Buolamwini e Gebu (2018) no estudo *Gender Shades*, apresentam taxas de erro mais altas ao identificar pessoas negras, especialmente mulheres. Esse viés está intrinsecamente ligado a bancos de dados pouco representativos, nos quais imagens de indivíduos brancos são super-representadas, enquanto imagens de pessoas negras são

sub-representadas, comprometendo a precisão dos resultados (Buolamwini; Gebru, 2018; Silva, 2022).

Os conjuntos de dados desbalanceados são um elemento central para a perpetuação de vieses algorítmicos. Buolamwini e Gebru (2018), por meio do estudo *Gender Shades*, demonstraram que a falta de representatividade demográfica nos dados de treinamento resultou em taxas de erro substancialmente mais altas para subgrupos interseccionais, como mulheres negras, em comparação com homens brancos.

A ausência de restrições explícitas voltadas à igualdade, conhecidas como "*fairness constraints*", tem contribuído para a perpetuação dessas desigualdades. Para Zafar *et al.* (2017), as *fairness constraints* ajudam a criar classificadores que equilibram precisão e justiça, limitando vieses nas margens de decisão. Chen e Sontag (2018) ressaltam que tais restrições não apenas melhoram a imparcialidade dos modelos, mas também preservam ou até aprimoram sua precisão geral.

Buolamwini e Gebru (2018) destacam que muitas empresas, ainda falham em adotar essas práticas, mesmo diante de evidências claras sobre a existência de desigualdades sistêmicas nos resultados. O estudo *Gender Shades* avaliou algoritmos de grandes empresas, como Microsoft, Face++ e IBM, constatando que os conjuntos de dados utilizados para treinamento eram altamente enviesados. As pesquisadoras identificaram uma predominância de imagens de homens brancos, enquanto mulheres negras estavam sub-representadas. Essa discrepância resultou em erros que ultrapassaram 34% para mulheres negras, comparados a menos de 1% para homens brancos. Embora essas empresas tenham implementado correções, os problemas persistem, ainda que em menor grau (Raji; Buolamwini, 2019).

O conceito de "desinteligência artificial" reforça a ironia de sistemas tecnicamente avançados que, em vez de solucionar problemas, acabam por reforçar preconceitos e desigualdades (Buolamwini; Gebru, 2018; Benjamin, 2019). Além disso, a dependência excessiva da automação e a crença equivocada na objetividade dos algoritmos obscurecem os danos causados por essas tecnologias (Oliveira *et al.*, 2022). Em contextos de segurança pública, essa confiança cega em sistemas enviesados pode resultar em injustiças graves, como prisões errôneas e vigilância discriminatória (Benjamin, 2019).

Para mitigar os impactos do racismo algorítmico e da “desinteligência artificial”, é indispensável investir em estratégias que promovam a transparência e a responsabilidade no uso da IA (Benjamin, 2019). Entre essas estratégias, destacam-se a criação de políticas públicas que exijam auditorias algorítmicas, a diversificação de bases de dados e a implementação de práticas que considerem os impactos sociais dessas tecnologias.

Alguns exemplos práticos ilustram a gravidade do viés algorítmico. Em 2015, o *Google Photos* rotulou erroneamente pessoas negras como “gorilas” (Grush, 2015). Já em 2017, o *Apple Face ID* apresentou dificuldades para reconhecer indivíduos asiáticos (Curtis, 2017). Outro caso emblemático foi o da *Amazon Rekognition*, que confundiu parlamentares americanos com criminosos, conforme denunciado pela American Civil Liberties Union (Snow, 2018). Esses incidentes reforçam a necessidade de dados mais representativos nos treinamentos de algoritmos.

Finalmente, é imprescindível a implementação de normas claras e supervisão rigorosa no uso de tecnologias de reconhecimento facial, principalmente em vigilância preditiva.

IMPACTO DO RECONHECIMENTO FACIAL NA SEGURANÇA PÚBLICA

Os mapas de crime, utilizados como ferramentas tradicionais no Brasil, remontam ao período imperial (Ferreira, 2011). Essas práticas de controle e vigilância possuem raízes profundas na história colonial, quando eram empregados para o controle populacional. Com o tempo, evoluíram para sistemas digitais contemporâneos, impulsionados pela digitalização de dados e pela popularização de técnicas algorítmicas (Silva, 2022). No entanto, esses sistemas frequentemente perpetuam o legado colonial, aplicando uma vigilância seletiva que prejudica grupos marginalizados (Souza; Cifuentes; Cosate, 2023).

A vigilância contemporânea, especialmente por meio de tecnologias como o reconhecimento facial, deve ser analisada em conexão com suas raízes históricas. Browne (2015, p. 8, tradução livre) argumenta que “a história da marcação a ferro no escravismo transatlântico antecipa os projetos de ‘classificação social’ nas práticas contemporâneas de vigilância”. Esse vínculo histórico ressalta como os sistemas de reconhecimento facial carregam vieses derivados

de discriminações passadas, reforçando a necessidade de uma análise crítica para garantir uma aplicação ética e justa da tecnologia (Silva, 2022).

As tecnologias de reconhecimento facial aplicadas à segurança pública, apresentam limitações que incluem que violam direitos e liberdades individuais, além de perpetuarem práticas discriminatórias contra minorias (Grother; Ngan; Hanaoka, 2019). Apesar de serem amplamente utilizadas em operações policiais, essas tecnologias frequentemente apresentam erros com consequências graves, minando sua confiabilidade e efetividade (Silva, 2021).

A expansão do uso do reconhecimento facial pelas forças de segurança pública tem levantado preocupações quanto a seus efeitos negativos (Silva, 2022). Como alerta Silva (2018), a aplicação indiscriminada desses sistemas pode reforçar preconceitos existentes, configurando uma forma de “desinteligência artificial” que disfarça abusos sob a aparência de neutralidade tecnológica.

Nesse sentido, Benjamin (2019) alerta para os riscos de confiar cegamente em tecnologias policiais baseadas em algoritmos. De forma semelhante, Noble (2020) destaca que o conceito de “desinteligência artificial” ilustra a ironia de como tecnologias avançadas podem reforçar problemas sociais como racismo e discriminação. Silva (2022), por sua vez, aponta que grande parte das pessoas identificadas por reconhecimento facial pertence a grupos minoritários, reforçando a percepção de que os algoritmos carregam tendências discriminatórias prejudiciais a essas populações.

A necessidade de dados equilibrados é destacada por Melzi *et al.* (2024), que apontam as bases desbalanceadas como a principal causa de erros nos sistemas de reconhecimento facial. Entre as soluções propostas, destacam-se o uso de dados sintéticos e a diversificação de bases de imagens para mitigar essas limitações. Além disso, Duarte e Ceia (2022) alertam que sistemas de vigilância e reconhecimento facial, quando mal calibrados, podem levar a prisões indevidas e à perpetuação de práticas policiais discriminatórias, comprometendo a confiança pública nas forças de segurança.

A ausência de regulamentação, agrava o problema. Embora a Lei Geral de Proteção de Dados – LGPD (Brasil, 2018), em seu artigo 4º, inciso III, a criação de uma legislação complementar para normatizar o uso de informações para fins de segurança pública, ainda não há um marco regulatório para o uso

de tecnologias de reconhecimento facial no Brasil. Isso expõe a população, sobretudo as minorias, a discriminações e violações de direitos fundamentais.

No contexto internacional, tecnologias de reconhecimento facial têm sido amplamente exploradas em operações de segurança. A China, por exemplo, utilizou essa tecnologia para monitorar o cumprimento dos *lockdowns* durante a pandemia de Covid-19; o Japão a empregou nos Jogos Olímpicos de 2021; e o Reino Unido a utilizou em eventos de grande porte, como a coroação do Rei Charles III. No entanto, cidades americanas como São Francisco, Boston e Oakland proibiram o uso dessa tecnologia pelas forças de segurança, citando preocupações com privacidade, precisão das ferramentas e potenciais violações de direitos civis, além da falta de regulamentação apropriada (Francisco, Hiurel e Rielli, 2020; Tajra, 2024).

No Brasil, o projeto “Panóptico”¹, desenvolvido pelo Centro de Estudos de Segurança e Cidadania (CESeC), identificou mais de 200 iniciativas municipais de reconhecimento facial, com investimentos elevados destinados às instituições de segurança pública. Esse cenário reforça a urgência de regulamentação (CESeC, 2024). O Projeto de Lei 3069/22, em tramitação no Congresso Nacional, visa regulamentar o uso de reconhecimento facial na segurança pública, autorizando sua aplicação para a identificação biométrica automatizada de indivíduos (Agência Câmara de Notícias, 2023). Entretanto, o projeto ainda está em fase de discussão e não foi aprovado.

A tabela 1 apresenta os dados observados pelo projeto “Panóptico”, destacando as regiões brasileiras que utilizam a tecnologia de reconhecimento facial na segurança pública em território nacional.

¹ Panóptico é um projeto do Centro de Estudo de Segurança e Cidadania (CESeC), com apoio da Open Society Foundations, que monitora a adoção da tecnologia de reconhecimento facial pelas instituições de segurança pública do Brasil.

Tabela 1 - Projetos ativos de reconhecimento facial por região no Brasil.

Região	Projetos Ativos	Pessoas potencialmente vigiadas	Estado com mais projetos ativos	Número de projetos ativos do estado	Estado da Região com mais pessoas potencialmente vigiadas	Número de pessoas potencialmente vigiadas no estado
Norte	22	7.179.474	Amazonas	5	Pará	2.512.651
Nordeste	38	19.091.563	Pernambuco	8	Bahia	7.717.717
Centro-Oeste	75	4.377.563	Goiás	65	Goiás	3.077.321
Sul	46	9.528.000	Rio Grande do Sul e Paraná	17	Rio Grande do Sul e Paraná	5.470.420
Sudeste	70	33.801.757	São Paulo	32	São Paulo	18.829.731

Fonte: Adaptado do PANÓPTICO, (2024)

Duarte (2022) observa que os estados continuam a investir na tecnologia sem restrições claras para sua aplicabilidade. A implementação dessas ferramentas sem medidas que mitiguem seus vieses tem gerado pressão para o banimento dessas tecnologias (Moreira e Peres, 2023).

OS BENEFÍCIOS DA TECNOLOGIA DE RECONHECIMENTO FACIAL NA SEGURANÇA PÚBLICA

O direito de se deslocar com segurança é garantido pela Constituição Federal de 1988. Segundo o artigo 144, cabe ao Estado, com o apoio dos cidadãos, preservar a ordem pública e proteger a sociedade. Dessa forma, órgãos como as polícias federais, civis e militares têm adotado tecnologias de reconhecimento facial para aprimorar a segurança pública (Almeida, 2022).

Embora existam debates éticos e legais sobre o uso dessa tecnologia, estudos apontam seus benefícios. De acordo com Galterio, Shavit e Hayajneh (2018), o reconhecimento facial tem potencial para localizar criminosos, identificar crianças desaparecidas e rastrear extremistas, especialmente em contextos de vigilância em massa.

No cenário internacional, a aplicação do reconhecimento facial tem avançado em conformidade com regulamentações que visam promover a transparência e o controle ético. Nos Estados Unidos, por exemplo, o projeto de lei *Facial Recognition Act* estabelece critérios rigorosos para o uso da tecnologia por forças policiais, exigindo mandados judiciais e impondo restrições ao uso

de dados biométricos para evitar abusos (Lieu, 2022). Já na União Europeia, as Diretrizes 05/2022 do Conselho Europeu de Proteção de Dados, buscam limitar o uso indiscriminado da tecnologia em espaços públicos, priorizando o equilíbrio entre segurança e proteção à privacidade (EDPB, 2022).

Além disso, destaca-se a possibilidade de integrar sistemas de reconhecimento facial a estratégias de segurança em aeroportos e fronteiras. Países como Brasil, Reino Unido e China utilizam essa tecnologia para identificar passageiros com documentação falsa ou vínculos com atividades ilícitas, otimizando o controle de fronteiras e reduzindo filas em pontos de verificação (Pisanu *et al.*, 2021).

No Brasil, a tecnologia já demonstrou resultados positivos. Durante o Carnaval de Salvador, em 2020, a identificação de foragidos foi facilitada por câmeras com reconhecimento facial, levando à detenção de 42 pessoas procuradas (Reis *et al.*, 2021; Pisanu *et al.*, 2021). No Japão, a aplicação da tecnologia nos Jogos Olímpicos de 2021 reforçou a segurança de atletas e espectadores (Silva, 2022). Esses resultados demonstram o potencial do reconhecimento facial como ferramenta estratégica na gestão de riscos e prevenção de crimes.

No entanto, a implementação responsável dessas tecnologias requer regulamentações que garantam a proteção de direitos fundamentais, evitando abusos e desigualdades. Projetos como o de Salvador e outros nos estados do Rio de Janeiro e Ceará podem servir de base para futuras normas que conciliem avanços tecnológicos com ética e privacidade (Reis *et al.*, 2021; Silva, 2022).

OS DESAFIOS DA TECNOLOGIA DE RECONHECIMENTO FACIAL

A tecnologia de reconhecimento facial no Brasil enfrenta inúmeros desafios que abrangem aspectos técnicos, éticos e regulatórios. A precisão desse tipo de reconhecimento está diretamente relacionada à qualidade dos bancos de dados utilizados, sendo influenciada por fatores como ângulo, iluminação, expressões faciais e resolução da imagem (Huber *et al.*, 2024).

Embora os aspectos técnicos do reconhecimento facial tenham sido amplamente estudados, os impactos demográficos — como idade, raça e gênero — ainda são negligenciados, comprometendo sua eficácia (Abdurrahim; Samad; Huddin, 2017). Esses vieses têm resultado em abordagens seletivas e prisões indevidas, alimentando a desconfiança pública (Buolamwini; Gebre,

2018). No contexto brasileiro, marcado por desigualdades raciais e elevado encarceramento, a adoção dessa tecnologia tende a reforçar práticas discriminatórias (Silva, 2022).

A Lei Geral de Proteção de Dados (LGPD) fornece diretrizes gerais para o tratamento de dados, mas não especifica critérios claros para o uso de tecnologias de vigilância por órgãos públicos (Cebrian *et al.*, 2024). Apesar disso, diversas cidades brasileiras já utilizam o reconhecimento facial como ferramenta para combater a violência urbana (Instituto Igarapé, 2020; CESeC, 2024), o que acentua o desafio de equilibrar segurança pública com a proteção dos direitos fundamentais.

Nos Estados Unidos, o movimento contra a violência policial, intensificado em 2020, pressionou empresas de tecnologia a interromperem contratos com forças policiais. Como resultado, Google, IBM e Amazon restringiram a comercialização de suas tecnologias de reconhecimento facial para segurança pública (Wiewiórowski, 2020). No entanto, conforme destaca Noble (2020), as elites tecnológicas do Vale do Silício evitam confrontar diretamente questões de racismo e discriminação, perpetuando os vieses presentes em seus produtos.

No Brasil, o Projeto de Lei 3.069/2022, em tramitação na Câmara dos Deputados, busca regulamentar o uso do reconhecimento facial pelas forças de segurança, proibindo sua aplicação isolada para prisões e exigindo validação humana. A proposta reflete a tentativa de equilibrar segurança pública e direitos individuais (Câmara dos Deputados, 2024). Já o PL 2.338/2023, que trata do Marco Regulatório da Inteligência Artificial, permite o uso da tecnologia em contextos de segurança, mas tem sido criticado por entidades civis por divergir de normas internacionais, como o *Artificial Intelligence Act* da União Europeia, que proíbe seu uso em espaços públicos devido aos altos riscos envolvidos (União Europeia, 2024).

Adicionalmente, países como China e Estados Unidos têm implementado normas para assegurar o uso ético da inteligência artificial, evidenciando uma tendência global para o tema (Pirró, 2024). Esse panorama internacional ressalta a urgência de o Brasil adotar práticas alinhadas às recomendações globais, garantindo uma utilização responsável e ética das tecnologias de reconhecimento facial.

Campanhas de banimento das tecnologias de reconhecimento facial têm ganhado destaque no Brasil e em outros países. Essas campanhas buscam conscientizar a sociedade sobre os riscos associados ao uso dessa tecnologia e pressionar por regulamentações mais rigorosas ou até mesmo por sua proibição, especialmente na segurança pública.

No Brasil, a campanha “Sai Da Minha Cara” mobilizou mais de 50 mandatos em 2022, tornando-se uma iniciativa articulada por diversas entidades contrárias ao uso dessas tecnologias (Silva, 2021). Outra ação importante, “Tire Meu Rosto da Sua Mira”, alerta sobre os impactos discriminatórios e os riscos aos direitos fundamentais associados ao uso da tecnologia pelas forças de segurança, contando com apoio de ONGs e especialistas (Azevedo *et al.*, 2023). Além disso, a campanha “Sem Câmera Na Minha Cara” questiona projetos municipais, como o de Recife, argumentando que essas ferramentas reforçam discriminações estruturais, especialmente contra pessoas negras (Silva, 2021).

No contexto internacional, campanhas como “*Con Mi Cara No*”, na Argentina, e “*No Nos Vean La Cara*”, no México, abordam questões similares, denunciando o aumento da vigilância estatal e as violações de direitos humanos (Villaroel, 2022; ADC, 2024). Essas iniciativas refletem as preocupações de organizações da sociedade civil e especialistas em tecnologia sobre o uso dessas ferramentas para vigilância em massa.

SOLUÇÕES PARA MITIGAR O VIÉS ALGORÍTMICO EM SISTEMAS DE RECONHECIMENTO FACIAL

A implementação de sistemas de reconhecimento facial na segurança pública tem gerado debates, especialmente devido aos vieses algorítmicos presentes na tecnologia. Essa problemática está associada ao conceito de “desinteligência artificial”, que denuncia a capacidade de sistemas tecnológicos avançados perpetuarem preconceitos e desigualdades em vez de mitigá-los (Benjamin, 2019; Noble, 2020).

Para lidar com esses desafios, algumas soluções têm sido propostas por pesquisadores e especialistas. Uma das abordagens mais relevantes é aumentar a representatividade dos conjuntos de dados utilizados para treinar os modelos de reconhecimento facial (Buolamwini; Gebru, 2018). Além disso, a realização de auditorias algorítmicas é uma das medidas propostas para

identificar e mitigar vieses nos sistemas de reconhecimento facial. Raji *et al.* (2020) enfatizam a importância de integrar essas auditorias ao ciclo de vida dos algoritmos, assegurando maior transparência no desenvolvimento e na aplicação dessas tecnologias.

Outra estratégia para a mitigação dos vieses são as auditorias sociotécnicas, que apresentam uma abordagem estruturada para avaliar o uso ético, técnico e legal das tecnologias de reconhecimento facial pelas forças policiais. A estrutura dessas auditorias é baseada em cinco pilares principais: conformidade legal, necessidade democrática, proteção de dados, não discriminação e liberdades fundamentais. Cada um desses pilares é analisado por meio de critérios objetivos que guiam a avaliação de práticas específicas (Radiya-Dixit, 2023).

Outra solução consiste no desenvolvimento de arquiteturas neurais especializadas. Segundo Madan *et al.* (2021), a criação de redes neurais com neurônios especializados pode aprimorar a capacidade dos sistemas de generalizar dados não vistos, reduzindo os vieses.

No campo regulatório, a formulação de diretrizes éticas e políticas específicas é igualmente essencial. Silva (2020) discute a urgência de regulamentações que enfrentem o racismo algorítmico e promovam justiça e igualdade na aplicação dessas tecnologias. As políticas públicas devem estabelecer limites claros para o uso de sistemas de reconhecimento facial, garantindo que sejam empregados de maneira ética e responsável.

A adoção de sistemas de reconhecimento facial na segurança pública exige um equilíbrio entre inovação tecnológica e princípios éticos. Medidas técnicas, como auditorias algorítmicas e a diversificação de dados, aliadas a regulamentações adequadas, são fundamentais para assegurar o uso responsável dessas ferramentas.

CONCLUSÃO

A análise realizada neste artigo evidencia que o uso de tecnologias de reconhecimento facial na segurança pública apresenta tanto oportunidades quanto riscos. Embora possam otimizar a identificação e apoiar ações mais eficazes, os desafios éticos e técnicos relacionados ao racismo algorítmico exigem

atenção urgente. A literatura mostra que esses sistemas podem aprofundar desigualdades e comprometer direitos fundamentais.

Entre as soluções, destacam-se as auditorias algorítmicas, a diversificação das bases de dados e a regulamentação específica como medidas essenciais para mitigar vieses e garantir transparência, equidade e responsabilidade. No Brasil, marcado por desigualdades raciais e sociais, a ausência de marcos regulatórios agrava os riscos de discriminação e vigilância excessiva, exigindo políticas públicas baseadas em evidências e alinhadas a boas práticas internacionais.

Acreditar na neutralidade algorítmica, sem considerar suas implicações sociais, compromete a credibilidade dessas tecnologias. Assim, é fundamental promover a integração entre inovação tecnológica e justiça social. Conclui-se que a eficácia do reconhecimento facial depende de sua aplicação ética, transparente e justa, assegurando segurança pública sem violar direitos fundamentais.

REFERÊNCIAS

ADC. Com Mi Cara No. **ADC - Por los Derechos Civiles**, 2024. Disponível em: <https://conmicarano.adc.org.ar/#>. Acesso em: 05 dez. 2024.

AGÊNCIA CÂMARA DE NOTÍCIAS. **Projeto regulamenta o uso de reconhecimento facial por forças de segurança pública**. Agência câmara de notícias, 2023. Disponível em: <https://www.camara.leg.br/noticias/946010-projeto-regulamenta-o-uso-de-reconhecimento-facial-por-forcas-de-seguranca-publica/>. Acesso em: 18 nov. 2024.

ALMEIDA, E. C. Os grandes irmãos: o uso de tecnologias de reconhecimento facial para persecução penal. **Revista Brasileira de Segurança Pública**, v. 16, n. 2, p. 264–283, 2022.

ARAÚJO, R. A.; CARDOSO, N. D.; PAULA, A. M. Regulação e uso do reconhecimento facial na segurança pública do Brasil. **Revista de Doutrina Jurídica**, v. 112, p. e021009–e021009, 2021. Disponível em: <https://www.indexlaw.org/index.php/doutrinajuridica/article/view/7896>. Acesso em: 10 dez. 2024.

AZEVEDO, Cynthia *et al.* "Tire meu rosto da sua mira": em busca do banimento de tecnologias de reconhecimento facial na segurança pública brasileira. In: CEIA, Eleonora; EDLER, Daniel (Orgs.). **Tecnologia, segurança e direitos: os usos e riscos de sistemas de reconhecimento facial no Brasil**. São Paulo: Konrad-Adenauer-Stiftung, 2023. Disponível em: <https://www.kas.de/pt/web/brasilien/einzeltitel/-/content/tecnologia-seguranca-e-direitos-1>. Acesso em: 04 dez. 2024.

BENJAMIN, Ruha. **Race after technology: abolitionist tools for the new Jim Code**. Cambridge: Polity Press, 2019. Disponível em: <https://mycours.es/ems2/files/2020/09/Ruha-Benjamin-%E2%80%93-Race-After-Technology-Introduction.pdf>. Acesso em: 25 set. 2024.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Institui a Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, 14 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 17 dez. 2024.

BROWNE, Simone. **Dark matters: on the surveillance of blackness**. Durham: Duke University Press, 2015. Disponível em: <http://digix.manoa.hawaii.edu/es305sp22/220405-07Surveillance/browne%20-%20Dark%20Matters%20Surveillance.pdf>. Acesso em: 27 nov. 2024.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, p. 1-15, 2018. Disponível em: <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>. Acesso em: 12 nov. 2024.

CEBRIAN, Fabiana *et al.* **Biometria e reconhecimento facial**. ANPD, 2024. Disponível em: <https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/>. Acesso em: 12 dez. 2024.

CESEC. O Panóptico: monitor do reconhecimento facial no brasil. Cesecc, 2024. Disponível em: <https://cesecseguranca.com.br/projeto/o-panoptico-monitor-do-reconhecimento-facial-no-brasil/>. Acesso em: 10 jul. 2024.

CHEN, Irene, JOHANSSON, Fredrik, SONTAG, David. Why is my classifier discriminatory? **ACM Digital Library**, 2018. Disponível em: <https://dl.acm.org/doi/pdf/10.5555/3327144.3327272>. Acesso em 15 jan. 2025.

CURTIS, Sophie. iPhone X racism row: Apple's Face ID fails to distinguish between Chinese users. **The Mirror**, 21 dez. 2017. Disponível em: <https://www.mirror.co.uk/tech/apple-accused-racism-after-face-11735152>. Acesso em: 26 nov. 2024.

DUARTE, Daniel; CEIA, Eleonora. **Tecnologia, segurança e direitos: os usos e riscos de sistemas de reconhecimento facial no Brasil**. Rio de Janeiro: Konrad-Adenauer-Stiftung, 2022. Disponível em: <https://nev.prp.usp.br/publicacao/tecnologia-seguranca-e-direitos-os-usos-e-riscos-de-sistemas-de-reconhecimento-facial-no-brasil-2/>. Acesso em: 26 nov. 2024.

FERREIRA, Ricardo Alexandre. *Crimes em comum: escravidão e liberdade sob a pena do Estado imperial brasileiro (1830-1888)*. São Paulo: Editora Unesp, 2011.

FRANCISCO, P. A. P.; HUREL, L. M.; RIELLI, M. M. Regulação do reconhecimento facial no setor público: avaliação de experiências internacionais. [s.l.]: Instituto Igarapé, e Data Privacy Brasil Research, 2020. Disponível em: <https://igarape.org.br/wp-content/uploads/2020/06/2020-06-09-Regula%C3%A7%C3%A3o-do-reconhecimento-facial-no-setor-p%C3%BAblico.pdf>. Acesso em: 17 nov. 2024.

GALTERIO, Mary Grace; SHAVIT, Simi Angelic; HAYAJNEH, Thaier. A review of facial biometrics security for smart devices. **Computers**, v. 7, n. 3, 2018. Disponível em: <https://www.mdpi.com/2073-431X/7/3/37>. Acesso em: 27 nov. 2024.

GROTHER, Patrick; NGAN, Mei; HANAOKA, Kayee. Face Recognition Vendor Test (FRVT) – Part 3: Demographic effects. **National Institute of Standards and Technology (NIST)**, 2019. Disponível em: <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>. Acesso em: 18 nov. 2024.

GRUSH, Loren. Google engineer apologizes after Photos app tags two Black people as gorillas. **The Verge**, 1 jul. 2015. Disponível em: <https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>. Acesso em: 10 dez. 2024.

HUBER, Marco *et al.* Bias and diversity in synthetic-based face recognition. **WACV**, 2024. Disponível em: <https://ieeexplore.ieee.org/document/10483704/authors#authors>. Acesso em: 05 dez. 2024.

INSTITUTO IGARAPÉ. Reconhecimento facial no Brasil. Instituto Igarapé, 2019. Disponível em: <https://igarape.org.br/infografico-reconhecimento-facial-no-brasil/>. Acesso em: 05 dez. 2024.

LI, Stan Z.; JAIN, Anil K. **Handbook of face recognition**. New York: Springer, 2011.

LIEU, Ted. Reprs Ted Lieu, Sheila Jackson Lee, Yvette Clarke, and Jimmy Gomez introduce bill to regulate law enforcement use of facial recognition technology. **Ted Lieu**, 2022. Disponível em: <https://lieu.house.gov/media-center/press-releases/reprs-ted-lieu-sheila-jackson-lee-yvette-clarke-and-jimmy-gomez-introduce>. Acesso em: 26 nov. 2024.

MADAN, Spandan *et al.* When and how CNNs generalize to out-of-distribution category-viewpoint combinations. **Computers and Society**, 17 nov. 2021. Disponível em: <https://arxiv.org/abs/2007.08032>. Acesso em: 17 DEZ. 2024.

MELZI, P. *et al.* FRCSyn-onGoing: benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems. **Information Fusion**, [S.l.], 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1566253524001003>. Acesso em: 8 dez. 2024.

MOREIRA, Horrara; PERES, Pedro. **Reconhecimento facial da segurança pública**: Por que banir? Heinrich-Böll-Stiftung, 2023. Disponível em: <https://br.boell.org/pt-br/2023/04/06/reconhecimento-facial-da-seguranca-publica-por-que-banir>. Acesso em: 17 nov. 2024

NOBLE, Safiya. **Algorithms of oppression: how search engines reinforce racism**. New York: NYU Press, 2020. Disponível em: https://safiyaunoble.com/wp-content/uploads/2020/09/Algorithms_Oppression_Introduction_Intro.pdf. Acesso em: 25 set. 2024.

OLIVEIRA, T. *et al.* Aspectos ético-jurídicos e tecnológicos do emprego de reconhecimento facial na segurança pública no Brasil. **Revista Tecnologia e Sociedade**, Curitiba, v. 18, n. 50, jan./mar. 2022. Disponível em: <https://periodicos.utfpr.edu.br/rts/article/view/12968>. Acesso em: 02 dez. 2024.

PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. (2015). Deep Face Recognition. **British Machine Vision Conference (BMVC)**, 2015. Disponível em: <https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/>. Acesso em: 29 dez. 2024.

PETRESCU, Rely. Face recognition as a biometric application. **Journal of Mechatronics and Robotics**, v. 3, s.n., 2019. Disponível em: <https://thescpub.com/abstract/jmrsp.2019.237.257>. Acesso em: 27 nov. 2024.

PHILLIPS, P. J.; YEROLOIMAS, D. M.; HILL, M. M.; O'TOOLE, A. J.; WEISS, D. J.; CHANG, J. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. **Proceedings of the National Academy of Sciences**, [S.l.], v. 115, n. 24, p. 6171–6176, 2018. Disponível em: <https://www.pnas.org/doi/full/10.1073/pnas.1721355115>. Acesso em: 5 dez. 2024.

PISANU, Gaspar *et al.* **Tecnologia de vigilância na América Latina**: feita no exterior, implantada em casa. Access Now, 2021. Disponível em: <https://www.accessnow.org/tecnologia-de-vigilancia-na-america-latina/>. Acesso em: 24 nov. 2024.

RADIYA-DIXIT, Elham. A sociotechnical audit: assessing police use of facial recognition. **Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, 2023. Disponível em: <https://dl.acm.org/doi/10.1145/3593013.3594084>. Acesso em: 22 jan. 2025.

RAJI, I.; BUOLAMWINI, J. Actionable auditing: investigating the impact of publicly naming biased performance results of commercial AI products. *In: AAAI/ACM Conference on AI, Ethics and Society*, 2019.

RAJI, Inioluwa Deborah; FRIED, Genevieve. **About Face: A Survey of Facial Recognition Evaluation**. AAAI 2020 Workshop on AI Evaluation, 2020. Disponível em: <https://arxiv.org/pdf/2102.00813>. Acesso em: 28 dez. 2024.

RAJI, I. D.; SMART, A.; WHITE, R. N.; LI, J.; DIXON, H.; GEBRU, T. Saving face: investigating the ethical concerns of facial recognition auditing. **Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society**, p. 145–151, 2020. Disponível em: <https://arxiv.org/abs/2001.00964>. Acesso em: 17 dez. 2024.

RAVIV, Shaun. The secret history of facial recognition. **Wired**, 21 jan. 2020. Disponível em: <https://www.wired.com/story/secret-history-facial-recognition/>. Acesso em: 19 set. 2024.

REIS, Carolina; ALMEIDA, Eduardo; DOURADO, Fernanda; SILVA, Rafael. Relatório sobre o uso de tecnologias de reconhecimento facial e câmeras de vigilância pela administração pública no Brasil. Brasília: Laboratório de Políticas Públicas e Internet, 2021. Disponível em: <https://lapin.org.br/2021/07/07/vigilancia-automatizada-uso-de-reconhecimento-facial-pela-administracao-publica-no-brasil/>. Acesso em: 26 nov. 2024.

SILVA, Tarcízio. *(Des)Inteligência artificial: como computadores não compreendem o mundo*. 2018. Disponível em: <https://tarciziosilva.com.br/blog/desinteligencia-artificial-como-computadores-nao-compreendem-o-mundo/>. Acesso em: 29 nov. 2024.

SILVA, Tarcízio. **Racismo algorítmico: inteligência artificial e discriminação nas redes digitais**. São Paulo: Sesc, 2022.

SINHA, Pawan; POGGIO, Tomaso. I think I know that face...**Nature**, vol.

384, p. 404, 1996. Disponível em: <https://www.nature.com/articles/384404a0>. Acesso em: 29 dez. 2024.

SINHA, Pawan. Qualitative Representations for Recognition. In: Bülthoff, H.H., Wallraven, C., Lee, SW., Poggio, T.A. (eds) *Biologically Motivated Computer Vision*. BMCV 2002. Lecture Notes in Computer Science, vol 2525. Springer, Berlin, Heidelberg. Disponível em: https://link.springer.com/chapter/10.1007/3-540-36181-2_25. Acesso em: 29 dez. 2024.

SILVA, Tarcízio. Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina. **Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)**, 2020. Disponível em: <https://abpnrevista.org.br/site/article/view/744>. Acesso em: 20 jun. 2024.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2015. Disponível em: https://www.cv-foundation.org/openaccess/content_cvpr_2015/app/1A_089.pdf. Acesso em: 29 dez. 2015.

SNOW, Jacob. Amazon's face recognition falsely matched 28 members of Congress with mugshots. **ACLU**, 26 jul. 2018. Disponível em: <https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28>. Acesso em: 26 nov. 2024.

SOUZA, Ramison; CIFUENTES, Tainá; COSATE, Tatiana. Racismo estrutural e punição no Brasil: o reconhecimento facial como nova ferramenta de seletividade estatal. **Criminologia Periférica**, 2023. Disponível em: <https://publicar.claec.org/index.php/editora/catalog/book/84>. Acesso em: 05 dez. 2024.

TAJRA, Alex. Veja como cada estado brasileiro utiliza o reconhecimento facial para fins policiais. **Consultor Jurídico**, 17 mai. 2024. Disponível em: <https://www.conjur.com.br/2024-mai-17/veja-como-cada-estado-usa-o-reconhecimento-facial-para-fins-policiais/>. Acesso em: 04 nov. 2024.

TUTEJA, Deepak *et al.* Facial Recognition System. **Journal of Basic and Applied Engineering Research**, 2014. Disponível em: https://krishisanskriti.org/vol_image/12Jun20150606054.pdf. Acesso em: 02 jan. 2025.

UNIÃO EUROPEIA. **Artificial Intelligence Act**. 2024. Disponível em: <https://artificialintelligenceact.eu/ai-act-explorer/>. Acesso em: 29 nov. 2024.

ABDURRAHIM, Salem Hamed, SAMAD, Salina Abdul, HUDDIN. Review on the effects of age, gender, and race demographics on automatic face recognition. **Vis Comput** 34, 1617–1630, 2017. Disponível em: <https://link.springer.com/article/10.1007/s00371-017-1428-z>. Acesso em: 29 nov. 2024.

VILLAROEL, Joselyn. **Basta de normalizar la vigilancia masiva en América Latina**. Access Now, 2022. Disponível em: <https://www.accessnow.org/basta-normalizar-vigilancia-masiva-america-latina/>. Acesso em: 12 dez. 2024.

VISHWANATHA C. R, *et al.*, Face Recognition and Identification Using Deep Learning. Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2023, pp. 1-5. Disponível em: <https://ieeexplore.ieee.org/document/10118154>. Acesso em: 29 dez. 2024.

ZAFAR, Muhammad Bilal; VALERA, Isabel; GUMMADI, Krishna P.; WROBEL, Stefan. Fairness constraints: Mechanisms for fair classification. Proceedings of the 20th **International Conference on Artificial Intelligence and Statistics (AISTATS)**, 2017. p. 962-970. Disponível em: <https://arxiv.org/pdf/1507.05259>. Acesso em: 15 jan. 2025.

WEN, Y.; HOLWEG, M. A phenomenological perspective on AI ethical failures: the case of facial recognition technology. **AI & Society**, v. 39, p. 1929–1946, 2024. Disponível em: <https://link.springer.com/article/10.1007/s00146-023-01648-7>. Acesso em: 10 dez. 2024.

WIEWIÓROWSKI, Wojciech. AI and facial recognition: challenges and opportunities. **European Data Protection Supervisor**, 21 fev. 2020. Disponível em: https://www.edps.europa.eu/press-publications/press-news/blog/ai-and-facial-recognition-challenges-and-opportunities_en. Acesso em: 30 nov. 2024.

RECONHECIMENTO FACIAL E REPRESENTATIVIDADE: A INFLUÊNCIA DAS VARIÁVEIS NA PRECISÃO DOS ALGORITMOS

Roberto Cezar Marcelino Neris
Universidade Tecnológica Federal do Paraná (UTFPR)

RESUMO

Este artigo explora como o desempenho dos algoritmos de reconhecimento facial é influenciado por fatores externos e características individuais de cada pessoa. A identificação desses elementos e a compreensão de seu impacto podem contribuir para aprimorar o desenvolvimento desses sistemas, aumentando sua precisão. Diferentes fatores afetam a eficácia do reconhecimento facial, como pose, iluminação, expressão e resolução das imagens, os quais são reconhecidos como desafios no campo do reconhecimento facial. Além disso, a influência das características demográficas da população em geral, como etnia, idade e gênero, é também um desafio que recebe atenção dos pesquisadores, no tocante ao desempenho dos algoritmos de reconhecimento facial.

Palavras-chave: reconhecimento facial; variáveis demográficas; inteligência artificial; algoritmos.

INTRODUÇÃO

O reconhecimento facial é amplamente utilizado em áreas como segurança pública, controle de acesso e autenticação digital. Trata-se de um sistema biométrico que analisa características faciais e as compara com modelos armazenados em um banco de dados (Li; Jain, 2009).

Apesar de sua crescente adoção, estudos indicam que a precisão dos sistemas de reconhecimento facial muda entre diferentes grupos populacionais, sendo sensível a fatores como idade, gênero e etnia (Buolamwini; Gebru, 2018). Essas disparidades podem resultar em falhas de identificação e contribuir para a reprodução de discriminações algorítmicas, especialmente em contextos de decisões automatizadas, como na vigilância e no policiamento (Grother; Ngan; Hanaoka, 2019).

A crescente adoção do reconhecimento facial por governos e empresas exige o desenvolvimento de tecnologias confiáveis e imparciais. No entanto, a limitada representatividade demográfica das bases de treinamento permanece como um dos principais desafios (Richardson; Schwarz; Sukumar, 2021).

Este artigo busca responder à seguinte questão: Como reduzir os vieses algorítmicos em sistemas de reconhecimento facial, considerando o impacto das variáveis demográficas? Para tanto, o estudo analisa a influência dessas variáveis no desempenho dos sistemas, identifica possíveis limitações das bases de dados, e propõe melhorias que possam contribuir para o aperfeiçoamento da tecnologia.

A pesquisa se justifica pela necessidade de mitigar os impactos negativos dos vieses algorítmicos relacionados a variáveis demográficas. A confiabilidade e imparcialidade dessa tecnologia são essenciais para prevenir injustiças e garantir um uso ético e equitativo em aplicações que afetam diretamente os direitos das pessoas.

Parte-se da hipótese de que a diversificação das bases de dados utilizadas no treinamento dos algoritmos, aliada à implementação de auditorias algorítmicas e diretrizes regulatórias, podem reduzir vieses em sistemas de reconhecimento facial. Com modelos mais representativos, espera-se aumentar a confiabilidade na aplicação da tecnologia.

Este estudo concentra-se na análise dos impactos dos vieses algorítmicos em sistemas de reconhecimento facial, utilizados principalmente na segurança pública. Embora existam aplicações em outras áreas, o foco da pesquisa recai sobre contextos em que erros podem acarretar a discriminação algorítmica. A revisão de literatura inclui estudos nacionais e internacionais, com ênfase nos desafios enfrentados no Brasil, onde desigualdades estruturais tendem a agravar os efeitos desses vieses.

Para estruturar a discussão, o artigo está organizado da seguinte forma: a primeira seção apresenta uma revisão teórica sobre o reconhecimento facial e suas aplicações; a segunda aborda os desafios relacionados às variáveis demográficas e à representatividade dos dados; a terceira discute estratégias para mitigação de vieses. Por fim, a conclusão reforça a necessidade de aprimorar os sistemas para torná-los mais justos e confiáveis.

REVISÃO TEÓRICA SOBRE RECONHECIMENTO FACIAL

O reconhecimento facial é uma tecnologia promissora, mas apresenta limitações na identificação de pessoas. Inúmeros estudos indicam que esses sistemas cometem erros ao interpretar características como gênero, idade, etnia (Waelen, 2023).

A tecnologia possibilita a identificação ou verificação de indivíduos a partir de suas características faciais e tem sido amplamente adotada em áreas como segurança pública, controle de acesso e autenticação digital (Gang *et al.*, 2009). A partir de imagens faciais, os sistemas extraem características que são convertidas em modelos digitais e comparadas a bases de dados (Smith; Miller, 2022).

Na segurança pública, a tecnologia é utilizada para monitoramento e identificação de suspeitos em ambientes públicos, auxiliando órgãos policiais e de inteligência (Osório, 2024). Empresas também empregam o reconhecimento facial para autenticar usuários em dispositivos móveis e plataformas digitais, garantindo maior segurança em transações online (Buolamwini; Gebu, 2018). No varejo, o uso da tecnologia contribui para a personalização da experiência do consumidor ao analisar expressões faciais e padrões comportamentais (Gang *et al.*, 2009).

A precisão do reconhecimento facial pode ser afetada por diferentes fatores, como iluminação, ângulo de captura, resolução da imagem e expressões faciais. Buolamwini e Gebru (2018) apontam que os algoritmos de reconhecimento facial, têm dificuldades em reconhecer com precisão mulheres negras, em comparação com homens brancos, devido à falta de representatividade demográfica nos conjuntos de dados de treinamento destes sistemas.

Uma possível solução para aumentar a representatividade demográfica é o uso de bases de dados diversificadas, que incluam imagens de diferentes idades, gêneros e etnias (Ferraz; Santos; Santos, 2023). Além disso, a combinação de dados sintéticos com imagens reais tem mostrado melhorias na precisão dos algoritmos sem comprometer a privacidade dos usuários (Melzi *et al.*, 2024).

Embora amplamente utilizada, a tecnologia enfrenta desafios relacionados à precisão, vieses algorítmicos, privacidade, segurança e questões éticas. O debate sobre seu uso levanta preocupações sobre a coleta de dados e a vigilância estatal, especialmente no contexto das forças policiais (Henschke, 2017).

VARIÁVEIS DO RECONHECIMENTO FACIAL

Convém observar que diferentes fatores influenciam a precisão dos sistemas de reconhecimento facial. A compreensão desses fatores é essencial para aprimorar o desempenho dos sistemas e mitigar vieses. A iluminação, o ângulo de captura, a resolução da imagem, a expressão facial e obstruções são variáveis que afetam diretamente a acurácia do reconhecimento facial. Estudos revelam que variações na iluminação reduzem a precisão dos algoritmos (Phillips *et al.*, 2005).

A iluminação é um dos maiores desafios da tecnologia, pois pode ocultar características essenciais para a identificação (Panceri *et al.*, 2015). Outros fatores, como idade, gênero e etnia, também impactam a precisão dos sistemas. Algoritmos treinados com bases de dados não representativas tendem a apresentar taxas de erro mais altas para determinados grupos (Buolamwini; Gebru, 2018).

Para Faria e Silva (2022), o uso de acessórios como óculos, chapéus, maquiagens e mudanças fisiológicas podem comprometer a precisão do reconhecimento, especialmente em contextos de segurança pública, onde a confiabilidade é fundamental. O ângulo da câmera e a qualidade da imagem, podem

ser controladas durante a captura, ao contrário das características pessoais, como etnia, gênero e faixa etária, que não são modificáveis pelos algoritmos (Abdurrahim; Samad; Huddin, 2018). A interação entre fatores como iluminação inadequada, poses variadas e características demográficas pode reduzir a acurácia dos algoritmos.

Variáveis controladas

As variáveis controladas são aquelas que podem ser ajustadas em ambientes ou aplicações específicas, e influenciam diretamente a precisão dos sistemas de reconhecimento facial. Essas variáveis devem ser consideradas para garantir a confiabilidade dos resultados.

Posição do rosto

A pose ou o ângulo de captura das imagens é um fator relevante no reconhecimento facial. Diferentes ângulos podem alterar as características faciais de uma pessoa, impactando a capacidade do sistema de reconhecimento facial em identificar corretamente um indivíduo (Phillips *et al.*, 2005).

A posição do rosto refere-se às diferentes angulações em relação à câmera durante a captura. Certas posições podem ocultar detalhes importantes, dificultando a identificação correta (Abdurrahim; Samad; Huddin, 2018). As imagens frontais ou de frente de uma pessoa, possuem mais informações para sua identificação do que imagens de perfil, que podem ocultar traços faciais. Em função disso, a precisão dos algoritmos tende a diminuir em imagens faciais que não são frontais. Pesquisadores têm desenvolvido soluções para minimizar esses efeitos, dividindo essas abordagens em três categorias: (1) extração de características tolerantes à pose, (2) *matching* baseado em visualizações e (3) transformação de pose, que utiliza modelos 3D para simular diferentes condições de captura (Zhang; Gao, 2009).

Efeitos da iluminação

A iluminação desempenha um papel importante na eficácia do reconhecimento facial. As variações na iluminação podem obscurecer detalhes faciais e dificultar a identificação pelos algoritmos (Abdurrahim; Samad; Huddin, 2018).

Este efeito se intensifica quando combinado com outros fatores, como pose, expressão facial e o uso de acessórios (Bhatt; Singh; Vatsa, 2015). De acordo com Bhatt, Singh e Vatsa (2015), existem abordagens passivas, como normalização fotométrica e equalização de histograma, e ativas, como o uso de câmeras infravermelhas ou imagens térmicas, que são menos suscetíveis às mudanças na iluminação ambiente.

Sistemas de reconhecimento facial são significativamente afetados pelas condições de iluminação. A taxa de erro aumenta em ambientes com iluminação desfavorável ou com sombras e reflexos (Damasceno, 2017).

Expressão facial

As mudanças nas expressões faciais podem alterar a aparência do rosto e, conseqüentemente, a precisão dos sistemas de reconhecimento facial (Bhatt; Singh; Vatsa, 2015). Determinadas expressões emocionais, como surpresa ou alegria, são mais facilmente identificadas, enquanto outras, como medo ou tristeza, podem reduzir a eficácia dos algoritmos (Silva Neto; Afonso; Souza, 2023). Além disso, microexpressões involuntárias introduzem desafios adicionais, pois são difíceis de distinguir dos movimentos naturais do rosto. A utilização de técnicas de aprendizado profundo tem demonstrado melhorar a precisão ao reduzir os efeitos negativos das variações expressivas (Koujan *et al.*, 2020).

Resolução de imagem

A resolução da imagem é determinante para a precisão do reconhecimento facial. As imagens de baixa resolução podem não capturar detalhes importantes, prejudicando a performance dos algoritmos (Melzi *et al.*, 2024). A redução na resolução afeta diretamente a capacidade de captura de texturas e contornos faciais, impactando a precisão do sistema (Knoche; Hörmann; Rigoll, 2021).

VARIÁVEIS NÃO CONTROLADAS

No reconhecimento facial, certas variáveis não podem ser controladas diretamente pelos algoritmos, mas afetam sua precisão. Entre elas, destacam-se a idade, o gênero e a etnia, fatores que alteram a aparência facial e podem dificultar a identificação.

Envelhecimento

O envelhecimento é um dos maiores desafios para o reconhecimento facial, uma vez que altera as características faciais ao longo do tempo. Fatores como exposição ao sol, mudanças no peso e o aparecimento de rugas impactam a precisão da identificação. A perda de desempenho aumenta à medida que a diferença de idade entre a imagem de referência e a imagem de teste cresce (Osman Ali *et al.*, 2015).

Idade

A idade afeta a precisão dos algoritmos de reconhecimento facial. O *National Institute of Standards and Technology* (NIST) identificou variações na taxa de acerto entre diferentes faixas etárias (Grother; Ngan; Hanaoka, 2019).

Gênero

As diferenças faciais entre homens e mulheres também influenciam a precisão do reconhecimento facial. Buolamwini e Gebre (2018), no estudo *Gender Shades*, evidenciam que sistemas de reconhecimento têm maior taxa de erro na identificação de mulheres, especialmente mulheres negras em comparação com homens brancos. A taxa de erro para homens brancos foi inferior a 1%, enquanto para mulheres negras alcançou 34,4%.

Etnia

Os algoritmos de reconhecimento facial têm desempenho superior na identificação de indivíduos pertencentes a grupos étnicos predominantes, nas bases de dados de treinamento. Assim, pessoas negras, asiáticas e indígenas enfrentam maior taxa de erro devido à sub-representação nesses dados (Krishnapriya *et al.*, 2020).

EFEITOS DAS VARIÁVEIS DEMOGRÁFICAS

Com a crescente adoção em inúmeras áreas, o reconhecimento facial ainda apresenta falhas associadas a variáveis demográficas, como idade, gênero e etnia, que influenciam diretamente o desempenho dos algoritmos. Fatores

como etnia, gênero e idade podem introduzir vieses nos algoritmos, comprometendo sua precisão e imparcialidade (Grother; Ngan; Hanaoka, 2019). Silva (2022), reforça que as taxas de erro relativas a falsos positivos podem ser de dez a cem vezes maiores para pessoas negras, asiáticas ou de povos originários.

Os efeitos da idade

A variação das características faciais ao longo do tempo impacta a precisão dos sistemas de reconhecimento facial. Os algoritmos apresentam menor desempenho na identificação de crianças e idosos, sendo mais precisos para adultos jovens (Drozdowski; Rathgeb; Busch, 2020). Grother, Ngan e Hanaoka (2019), corroboram essa análise, demonstrando que a taxa de erro em crianças, pode ser até quatro vezes maior do que em adultos jovens, o que evidencia a necessidade de modelos mais representativos para diferentes faixas etárias.

Estudos mostram que indivíduos mais jovens tendem a apresentar maiores taxas de erro. Givens *et al.* (2004), utilizando o banco de dados FERET, constataram que algoritmos de reconhecimento facial têm mais dificuldades em reconhecer indivíduos mais jovens, independentemente do tipo de algoritmo utilizado.

Nesse contexto, o reconhecimento facial em crianças e jovens representa um desafio particular, já que suas feições mudam rapidamente durante o crescimento. Bahmani e Schuckers (2022), utilizando o conjunto de dados Young Face Aging (YFA), observaram uma queda nas taxas de correspondência à medida que aumentava o intervalo de tempo entre as imagens, mesmo em períodos curtos de seis meses. Entretanto, ao utilizarem o sistema *MagFace*, observaram taxas de aceitação verdadeira de 98,3% e 94,9% para intervalos de 6 e 36 meses, respectivamente, sugerindo viabilidade do reconhecimento facial infantil em intervalos de até três anos.

O impacto do envelhecimento também tem sido amplamente investigado. Albiero *et al.* (2019) analisaram a precisão de três modelos baseados em redes neurais convolucionais profundas (CNNs) em diferentes faixas etárias (16–29, 30–49 e 50–70 anos). Contrariando estudos anteriores, os resultados mostraram que a precisão diminui entre indivíduos mais velhos, sendo maior entre os mais jovens.

O efeito do gênero

A precisão dos algoritmos de reconhecimento facial muda conforme o gênero. Pesquisas indicam que esses sistemas são geralmente mais precisos para homens do que para mulheres, especialmente para mulheres negras (Grother; Ngan; Hanaoka, 2019). As taxas de erro são mais elevadas para mulheres negras, pois decorrem da sub-representação desses grupos nos conjuntos de dados de treinamento. Adicionalmente, fatores como morfologia facial, iluminação e acessórios (maquiagem, penteados) influenciam os resultados (Buolamwini; Gebu, 2018). Albiero *et al.* (2021), analisaram essas variáveis e concluíram que elementos como penteados podem cobrir partes do rosto, comprometendo a precisão. De forma semelhante, Bhatta *et al.* (2022) apontam que alterações estéticas comuns entre mulheres impactam o desempenho dos algoritmos.

Os vieses de gênero também estão presentes no reconhecimento de expressões faciais. Dominguez-Catena, Paternain e Galar (2022) demonstraram que a representação desequilibrada de gêneros nos bancos de dados influencia negativamente a precisão dos modelos, reforçando a importância da diversidade nos dados de treinamento.

O efeito da etnia

Pesquisas apontam que algoritmos de reconhecimento facial apresentam menor precisão na identificação de pessoas negras e de outras minorias raciais em comparação a pessoas brancas. Isso se deve à baixa representatividade desses grupos nos conjuntos de dados de treinamento e às variações fenotípicas que não são adequadamente processadas pelos sistemas (Silva, 2022).

Para Silva (2022) a combinação entre vieses algorítmicos e estruturas sociais discriminatórias, resulta no fenômeno chamado “racismo algorítmico”. Buolamwini e Gebu (2018), demonstraram que algoritmos apresentam maiores taxas de erro na identificação de mulheres negras e asiáticas, evidenciando o impacto da sub-representação demográfica desses grupos. Yucer *et al.* (2021), propuseram uma abordagem baseada em atributos fenotípicos, ou seja, que determina a aparência do indivíduo, como altura, cor dos olhos, cor do cabelo e outros, para identificar vieses raciais. Kolla e Savadamuthu (2022) investigaram a distribuição racial nos conjuntos de dados e concluíram que, mesmo com

uma distribuição equilibrada, os vieses persistem indicando que a equidade na amostragem, por si só, não garante justiça algorítmica.

Outro fenômeno observado é o *other-race effect*, que descreve a tendência de indivíduos reconhecerem rostos de sua própria etnia com maior precisão (Cavazos *et al.*, 2019). Todavia, efeito também se aplica a sistemas automatizados, que apresentam dificuldades em reconhecer rostos de grupos sub-representados nos dados de treinamento (Yucer *et al.*, 2023).

No contexto da segurança dos sistemas de reconhecimento facial, o *other-race effect* influencia a suscetibilidade a ataques de *morphing*, técnica esta que cria imagens faciais manipuladas para enganar os sistemas (Mallick *et al.*, 2022). Desse modo, tanto humanos quanto algoritmos são suscetíveis a erros com rostos de outras etnias, o que favorece o aumento dos riscos a ataques desse tipo.

BASE DE DADOS DEMOGRÁFICAS

Estudos demonstram que a falta de representatividade nos conjuntos de dados pode levar a vieses algorítmicos, afetando desproporcionalmente grupos sub-representados (Buolamwini; Gebru, 2018). Assim, garantir variabilidade demográfica, incluindo etnia, gênero e faixa etária, é fundamental para assegurar a confiabilidade dos sistemas de reconhecimento facial.

Pesquisas nessa área, tem se beneficiado de bases de dados demográficas que auxiliam no treinamento e avaliação dos algoritmos, reunindo atributos como idade, gênero, etnia, cor dos olhos e formato do rosto. Raji e Fried (2021), analisaram mais de 100 conjuntos de dados criados entre 1976 e 2019, totalizando 145 milhões de imagens de 17 milhões de indivíduos, e observaram que esses repositórios refletem mudanças políticas, tecnológicas e sociais, o que reforça a necessidade de maior transparência sobre sua composição.

Bharati *et al.* (2017) examinaram o banco *Multi-Demographic Retouched Faces* (MDRF) e identificaram que a precisão na detecção de retoques varia conforme os atributos demográficos das imagens de treinamento. Queiroz (2023) demonstrou que, no contexto brasileiro, vieses algorítmicos relacionados à população negra podem perpetuar desigualdades estruturais.

Nos últimos anos, pesquisadores têm se dedicado à ampliação e diversificação dessas bases. Iniciativas como o *Diversity in Faces* (Merler *et al.*, 2019) e o *CASIA-Face-Africa* (Muhammad *et al.*, 2021) visam aumentar a representatividade e reduzir os vieses dos modelos. Além disso, o uso de dados sintéticos, como os gerados pelo *DigiFace-1M* tem se mostrado promissor para equilibrar a distribuição de atributos faciais (Bae *et al.*, 2022).

A Tabela 1 apresenta alguns dos principais conjuntos de dados utilizados na pesquisa e desenvolvimento das variáveis demográficas.

Tabela 1. Pesquisas que lidaram com variáveis demográficas.

Base de Dados	Número de Imagens	Número de Indivíduos	Características
FG-NET	1.002	82	Idades variadas (de bebês a 69 anos)
VGGFace2	3,31 milhões	9.131	Variações em pose, idade, iluminação e etnia
CASIA-Face-Africa	38.546	1.183	Foco em indivíduos africanos
FERET	1.000+	N/D	Condições controladas, incluindo variações faciais
DigiFace-1M	1 milhão	Sintético	Dados sintéticos para reduzir vieses algorítmico
Diversity in Faces	1 milhão	N/D	Enfatiza diversidade facial
Labeled Faces in the Wild (LFW)	13.000	5.749	Coletado da internet, condições não controladas
MegaFace	1 milhão+	690.000	Teste de reconhecimento facial em larga escala
CelebA	200.000+	10.000+	Contém 40 atributos faciais anotados por imagem

Fonte: Elaborado pelo autor.

RECOMENDAÇÕES PARA A DIVERSIFICAÇÃO E MELHORIA DAS BASES DE DADOS DE RECONHECIMENTO FACIAL

A construção de bases de dados representativas é crucial para melhorar o desempenho e a justiça algorítmica. Buolamwini e Gebre (2018) demonstraram que a menor precisão na identificação de mulheres negras decorre da sub-representação dessa demografia nos dados de treinamento.

Melzi *et al.* (2024) propõem o uso de dados sintéticos gerados por Inteligência Artificial para ampliar a diversidade étnica e social, corrigindo lacunas de representatividade. A combinação entre dados reais e sintéticos oferece flexibilidade e autenticidade, resultando em melhores desempenhos dos algoritmos. Nesse sentido, Kong *et al.* (2019) destacam que o uso de aprendizado de máquina pode auxiliar na mitigação de vieses por meio da diversificação das bases de dados.

Para Queiroz (2023) a baixa diversidade nos dados de treinamento compromete a acurácia dos modelos para determinados grupos demográficos. Enquanto Grother, Ngan e Hanaoka (2019) reforçam a importância de equilibrar as faixas etárias nas bases de dados utilizadas para treinar aos algoritmos. Assim, tanto a idade dos indivíduos quanto o tamanho das bases de dados, influenciam diretamente na precisão dos sistemas.

Entre as soluções propostas para melhorar as bases de dados de treinamento dos algoritmos de reconhecimento facial, destacam-se as auditorias algorítmicas e o aprimoramento das bases de dados, com maior inserção de grupos demográficos (Raji *et al.*, 2021). A combinação dessas medidas é imperativa para criar sistemas mais confiáveis e justos, principalmente, quando utilizado na segurança para policiamento preditivo.

OS DESAFIOS NA REPRESENTATIVIDADE DAS BASES DE DADOS

A literatura existente aponta que um dos principais desafios enfrentados pelos pesquisadores da área é a ausência de bases de dados com representatividade adequada em termos de etnia, gênero e idade. Buolamwini e Gebre (2018) destacam que algoritmos treinados com dados desequilibrados tendem a apresentar taxas de erro mais elevadas na identificação de pessoas negras, principalmente mulheres negras. As variáveis controláveis, como iluminação, posição facial e expressão e não controláveis, como idade, etnia e gênero, impactam diretamente o desempenho dos sistemas (Li; Jain, 2011).

A base *CASIA-Face-Africa*, por exemplo, reúne 38.546 imagens de 1.183 indivíduos africanos, com anotações manuais de 68 pontos-chave faciais em diferentes condições de iluminação (Muhammad *et al.*, 2021), sendo um recurso relevante para o estudo de vieses raciais. Estudos também mostram que a

precisão dos algoritmos tende a diminuir com o envelhecimento dos indivíduos (Kemelmacher-Shlizerman *et al.* 2016).

A sub-representação de grupos étnicos minoritários nos dados de treinamento, contribui para o viés racial dos algoritmos (Silva, 2022). Portanto, a principal dificuldade enfrentada ainda reside na deficiência das bases de dados suficientemente representativas em termos demográficos.

CONCLUSÃO

O reconhecimento facial tem se mostrado uma ferramenta poderosa em diversas áreas, desde a segurança pública até a personalização de serviços. A implementação da tecnologia de reconhecimento facial, enfrenta desafios relacionados à representatividade demográfica, pois os sistemas dessa tecnologia tendem a apresentar vieses que afetam desproporcionalmente grupos como pessoas negras, asiáticas e mulheres.

Estudos demonstram que as variáveis demográficas influenciam diretamente a precisão dos algoritmos de reconhecimento facial, impactando a eficiência da tecnologia e colocando em riscos esses grupos. Esse cenário evidencia a necessidade do desenvolvimento de bases de dados de treinamento algorítmico mais representativas que considere a diversidade populacional e reduza as desigualdades de precisão entre diferentes grupos.

A acurácia dos sistemas é geralmente maior para homens do que para mulheres. Além disso, a diferença na taxa de reconhecimento entre gêneros diminui com o avanço da idade, sendo indivíduos mais velhos reconhecidos com maior facilidade em comparação aos mais jovens.

Um dos principais desafios para uma análise abrangente dos impactos das variáveis demográficas é a ausência de bases de dados suficientemente amplas e diversificadas. A necessidade de conjuntos de dados balanceados, que incluam diferentes faixas etárias, gêneros e etnias, é fundamental para garantir sistemas mais representativos e precisos.

Logo, a combinação de dados sintéticos e reais tem se mostrado eficaz para reduzir vieses. A transparência nos critérios de classificação de gênero é essencial para evitar discriminações, especialmente em relação a identidades trans e não binárias. Além do mais, desenvolvedores de algoritmos devem

considerar a complexidade das identidades de gênero e etnia para criar sistemas mais precisos, especialmente no contexto da segurança pública.

Por fim, o desenvolvimento de conjuntos de dados mais representativos e a realização de auditorias são também alguns dos passos importantes para superar os desafios da tecnologia de reconhecimento facial enfrenta tanto da comunidade científica quanto da sociedade.

REFERÊNCIAS

ABDURRAHIM, S. H.; SAMAD, S. A.; HUDDIN, A. B. Review on the effects of age, gender, and race demographics on automatic face recognition. **Visual Computer**, v. 34, p. 1617–1630, 2018. Disponível em: <https://doi.org/10.1007/s00371-017-1428-z>. Acesso em: 19 mar. 2025.

ALBIERO, Vítor; BOWYER, Kevin W.; VANGARA, Kushal; KING, Michael C. Does face recognition accuracy get better with age? Deep face matchers say no. **arXiv preprint**, arXiv:1911.06396, 2019. Disponível em: <https://arxiv.org/abs/1911.06396>. Acesso em: 3 mar. 2025.

ALBIERO, Vítor; ZHANG, Kai; KING, Michael C.; BOWYER, Kevin W. Gendered differences in face recognition accuracy explained by hairstyles, makeup, and facial morphology. **arXiv preprint**, arXiv:2112.14656, 2021. Disponível em: <https://arxiv.org/abs/2112.14656>. Acesso em: 3 mar. 2025.

BAE, Gwangbin *et al.* DigiFace-1M: 1 million digital face images for face recognition. **arXiv preprint**, arXiv:2210.02579, 2022. Disponível em: <https://arxiv.org/abs/2210.02579>. Acesso em: 5 mar. 2025

BAHMANI, Keivan; SCHUCKERS, Stephanie. Face recognition in children: a longitudinal study. **arXiv preprint**, arXiv:2204.01760, 2022. Disponível em: <https://arxiv.org/abs/2204.01760>. Acesso em: 3 mar. 2025.

BHARATI, Aparna *et al.* Demography-based facial retouching detection using subclass supervised sparse autoencoder. **arXiv preprint**, arXiv:1709.07598, 2017. Disponível em: <https://arxiv.org/abs/1709.07598>. Acesso em: 5 mar. 2025.

BHATTA, Aman *et al.* The gender gap in face recognition accuracy is a hairy problem. **arXiv preprint**, arXiv:2206.04867, 2022. Disponível em: <https://arxiv.org/abs/2206.04867>. Acesso em: 4 mar. 2025.

BHATT, H. S.; SINGH, R.; VATSA, M. Covariates of face recognition. Technical report submitted to **Delhi Institutional Repository**, 2015. Disponível em: <https://repository.iiitd.edu.in/jspui/bitstream/handle/123456789/214/IIITD-TR-2015-002.pdf?isAllowed=y&sequence=1>. Acesso em: 19 mar. 2025.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. *In: Conference on Fairness, Accountability, and Transparency*. Proceedings of Machine Learning Research, v. 81, p. 1-15, 2018. Disponível em: <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>. Acesso em: 12 nov. 2024.

CAVAZOS, Jonathan *et al.* Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *In: IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020. Anais [...]. [S.l.]: IEEE, 2020. p. 2144–2153. Disponível em: <https://doi.org/10.1109/WACV45572.2020.9093435>. Acesso em: 5 mar. 2025.

CHEN, B.-C.; CHEN, C.-S.; HSU, W. H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. **IEEE Transactions on Multimedia**, v. 17, n. 6, p. 804–815, 2015. Disponível em: <https://ieeexplore.ieee.org/document/7080893>. Acesso em: 19 mar. 2025.

DAMASCENO, Giselly Soares de Sousa. Reconhecimento facial com variações de iluminação utilizando PCA e modificações da DCT associadas aos classificadores GMM, Naïve Bayes e K-NN. 2017. 124 f. Dissertação (Mestrado em Ciência da Computação) – **Universidade Estadual do Ceará, Fortaleza**, 2017. Disponível em: <https://www.uece.br/wp-content/uploads/sites/51/2020/02/GISELLY-SOARES.pdf>. Acesso em: 19 mar. 2025.

DOMINGUEZ-CATENA, Iris; PATERNAIN, Daniel; GALAR, Mikel. Gender stereotyping impact in facial expression recognition. **arXiv preprint**, arXiv:2210.05332, 2022. Disponível em: <https://arxiv.org/abs/2210.05332>. Acesso em: 4 mar. 2025.

DROZDOWSKI, Patrick *et al.* Demographic bias in biometrics: A survey on an emerging challenge. **IEEE Transactions on Technology and Society**, v. 1, n. 2, p. 89–103, 2020. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9086771>. Acesso em: 3 mar. 2025.

FARIA, Renan Barboza de; SILVA, Renatha Gonçalves da Mota e. Breves considerações acerca da utilização do reconhecimento facial como instrumento de segurança pública e perseguição. **Boletim IBCCRIM**, 2022. Disponível em: https://publicacoes.ibccrim.org.br/index.php/boletim_1993/article/view/1561. Acesso em: 2 fev. 2025.

FERRAZ, H. G.; SANTOS, C. V. N.; SANTOS, R. P. N. Reconhecimento facial: os aspectos técnicos e controversos da tecnologia na sociedade e no direito processual penal. **Universidade Federal Fluminense**, 2023. Disponível em: <https://educapes.capes.gov.br/handle/capes/758606>. Acesso em: 12 mar. 2025.

GANG, L. *et al.* Geometric feature-based facial expression recognition using multiclass support vector machines. *In: IEEE International Conference on Granular Computing*, 2009. p. 318–321. Disponível em: <https://ieeexplore.ieee.org/document/5255147>. Acesso em: 12 mar. 2025.

GIVENS, G.; BEVERIDGE, J. R.; DRAPER, B. A.; GROTH, P.; PHILLIPS, P. J. How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. *In: IEEE Computer Society Conference On Computer Vision And Pattern Recognition (Cvpr)*, 2004. Proceedings... [S. l.]: IEEE, 2004. v. 2, p. Disponível em: <https://doi.org/10.1109/CVPR.2004.1315185>. Acesso em: 13 maio 2025.

GROTH, Patrick; NGAN, Mei; HANAOKA, Kayee. Face Recognition Vendor Test (FRVT) – Part 3: Demographic Effects. Gaithersburg: **National Institute of Standards and Technology (NIST)**, 2019. Disponível em: <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>. Acesso em: 28 fev. 2025.

HAGHIGHAT, M.; ABDEL-MOTALEB, M. Low resolution face recognition in surveillance systems using discriminant correlation analysis. *In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE & GESTURE RECOGNITION (FG)*, 12., 2017. Anais... [S. l.]: IEEE, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7961841>. Acesso em: 19 mar. 2025.

HENSCHKE, A. Ethics in an age of surveillance. Cambridge: **Cambridge University Press**, 2017. Disponível em: <http://bit.ly/3ZTFbYU>. Acesso em: 12 mar. 2025.

KEMELMACHER-SHLIZERMAN, Ira, *et al.* The MegaFace Benchmark: 1 million faces for recognition at scale. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, p. 4873–4882, 2016. Disponível em: <https://arxiv.org/pdf/1512.00596.pdf>. Acesso em: 18 de nov. 2023.

KOUJAN, Mohammad Rami *et al.* Real-time facial expression recognition “in the wild” by disentangling 3D expression from identity. **arXiv preprint**, arXiv:2005.05509, 2020. Disponível em: <https://arxiv.org/abs/2005.05509>. Acesso em: 25 fev. 2025.

- KONG, X. *et al.* Mitigate Bias in Face Recognition using Skewness-Aware Reinforcement Learning. **arXiv preprint**, arXiv:1901.08655, 2019. Disponível em: <https://arxiv.org/abs/1901.08655>. Acesso em: 14 mar. 2025.
- KOLLA, Manideep; SAVADAMUTHU, Aravindh. The impact of racial distribution in training data on face recognition bias: a closer look. **arXiv preprint**, arXiv:2211.14498, 2022. Disponível em: <https://arxiv.org/abs/2211.14498>. Acesso em: 4 mar. 2025.
- KNOCHE, Martin; HÖRMANN, Stefan; RIGOLL, Gerhard. Susceptibility to image resolution in face recognition and training strategies. **arXiv preprint**, arXiv:2107.03769, 2021. Disponível em: <https://arxiv.org/abs/2107.03769>. Acesso em: 25 fev. 2025.
- KRISHNAPRIYA, K. S.; ALBIERO, V.; VANGARA, K.; KING, M. C.; BOWYER, K. W. Issues Related to Face Recognition Accuracy Variation: A Case Study on Race. In: **IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**, 2020. p. 1-10. Disponível em: <https://arxiv.org/abs/2005.01014>. Acesso em: 7 mar. 2025.
- LI, Stan Z.; JAIN, Anil K. Handbook of Face Recognition. p. 301-327. New York: **Springer**, 2011. Disponível em: <https://link.springer.com/book/10.1007/978-0-85729-932-1>. Acesso em: 20 abr. 2025.
- LI, Stan Z.; JAIN, Anil (ed.). Encyclopedia of biometrics. [S. l.]: **Springer**, 2009. Disponível em: <https://link.springer.com/referencework/10.1007/978-0-387-73003-5>. Acesso em: 3 mar. 2025.
- LIU, Ziwei; LUO, Ping; WANG, Xiaogang; TANG, Xiaoou. Large-scale CelebFaces Attributes (CelebA) Dataset. **The Chinese University of Hong Kong**, 2021. Disponível em: <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>. Acesso em: 19 mar. 2025
- LYONS, Michael J. Excavating AI: Re-excavated. Debunking a fallacious account of the JAFFE dataset. 2020. **Pré-print**. Disponível em: <https://arxiv.org/abs/2009.01215>. Acesso em: 19 mar. 2025.
- MALLICK, Snipta; JECKELN, Geraldine; PARDE, Connor J.; CASTILLO, Carlos D.; O'TOOLE, Alice J. The influence of the other-race effect on susceptibility to face morphing attacks. **arXiv preprint**, arXiv:2204.12591, 2022. Disponível em: <https://arxiv.org/abs/2204.12591>. Acesso em: 4 mar. 2025.
- MELZI, P. *et al.* FRCSyn Challenge at WACV 2024: Face Recognition Challenge in the Era of Synthetic Data. Proceedings of the IEEE/CVF **Winter Conference on Applications of Computer Vision Workshops**, 2024. Disponível em: <https://frcsyn.github.io/>. Acesso em: 12 mar. 2025.
- MERLER, Michele; RATHA, Nalini; FERRARA, Michael; BOURDEV, Lubomir; HOWARD, Andrew; KOCHER, Patrick; YANG, Jing. Diversity in Faces. **IBM Journal of Research and Development**, v. 63, n. 4/5, p. 3:1-3:10, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8851349>. Acesso em: 7 mar. 2025.
- MUHAMMAD, Jawad; WANG, Yunlong; WANG, Caiyong; ZHANG, Kunbo; SUN, Zhenan. CASIA-Face-Africa: **A Large-scale African Face Image Database**. 2021. Disponível em: <https://arxiv.org/abs/2105.03632>. Acesso em: 20 mar. 2025
- OSÓRIO, F. Avanços e desafios da implantação da tecnologia de reconhecimento facial. **Jornal da USP**, 2024. Disponível em: <https://jornal.usp.br/atualidades/avancos-e-desafios-da-implantacao-da-tecnologia-de-reconhecimento-facial/>. Acesso em: 12 mar. 2025.
- OSMAN ALI, A. S. *et al.* Age-invariant face recognition using triangle geometric features. In: International Conference on Advances in Biometrics. Berlin: **Springer**, 2013. p. 71-78. (Lecture Notes in Computer Science, v. 5558). Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-40567-9_8. Acesso em: 19 mar. 2025.

PANCERI, João Antonio Campos *et al.* Reconhecimento facial baseado em HOG e PCA: uma comparação quanto à invariância à iluminação. **Revista Ifes Ciência**, v. 1, n. 1, 2015. Disponível em: <https://ojs.ifes.edu.br/index.php/ric/article/view/236>. Acesso em: 2 fev. 2025.

PHILLIPS, P. J. *et al.* An Introduction to the Face Recognition Grand Challenge. In: **IEEE Computer Society Conference On Computer Vision And Pattern Recognition (CVPR)**, 2005. p. 947-954. Disponível em: <https://doi.org/10.1109/CVPR.2005.268>. Acesso em: 14 mar. 2025.

QUEIROZ, Guilherme Matheus. A inteligência artificial e o reconhecimento facial: impactos à população negra no Brasil. 2023. Dissertação (Mestrado Profissional em Direito, Justiça e Desenvolvimento) – **Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa**, São Paulo, 2023. Disponível em: <https://repositorio.idp.edu.br//handle/123456789/4800>. Acesso em: 5 mar. 2025.

RAJI, Inioluwa Deborah; FRIED, Genevieve. About face: A survey of facial recognition evaluation. **arXiv preprint**, arXiv:2102.00813, 2021. Disponível em: <https://arxiv.org/abs/2102.00813>. Acesso em: 5 mar. 2025.

RICHARDSON, J.; SCHWARZ, C.; SUKUMAR, S. Bias in Face Recognition and Ways Forward. **Journal of Artificial Intelligence Research**, v. 70, p. 1417-1441, 2021. Disponível em: <https://doi.org/10.1613/jair.112234>. Acesso em: 7 mar. 2025.

SILVA NETO, Júlio Alves da; AFONSO, Stevam Lopes Alves; SOUZA, Wânia Cristina de. A utilização da imitação facial em tarefas de reconhecimento de expressões emocionais. **Psicologia: Ciência e Profissão**, v. 43, e249386, 2023. Disponível em: https://docs.bvsalud.org/biblioref/2023/08/1422400/a-utilizacao-da-imitacao-facial-em-tarefas-de-reconhecimento-d_QDwrHBt.pdf. Acesso em: 25 fev. 2025.

SILVA, Tarcízio. Racismo algorítmico: inteligência artificial e discriminação nas redes digitais. São Paulo: **Sesc**, 2022. Disponível em: <https://periodicos.ufac.br/index.php/RFIR/article/download/6482/4317/25412>. Acesso em: 13 maio 2025

SMITH, M.; MILLER, S. The ethical application of biometric facial recognition technology. **AI & Soc**, v. 37, p. 167-175, 2022. Disponível em: <https://doi.org/10.1007/s00146-021-01199-9>. Acesso em: 12 mar. 2025.

WAELEN, R. A. The struggle for recognition in the age of facial recognition technology. **AI Ethics**, v. 3, p. 215–222, 2023. Disponível em: <https://doi.org/10.1007/s43681-022-00146-8>. Acesso em: 12 mar. 2025.

WANG, Mei; DENG, Weihong; HU, Jiani; TAO, Xunqiang; HUANG, Yaohai. Racial Faces in-the-Wild: **Reducing Racial Bias by Information Maximization Adaptation Network**. 2023. Disponível em: <https://arxiv.org/pdf/1812.00194>. Acesso em: 19 mar. 2025.

YUCER, Seyma; TEKTAS, Furkan; AL MOUBAYED, Noura; BRECKON, Toby P. Measuring hidden bias within face recognition via racial phenotypes. **arXiv preprint**, arXiv:2110.09839, 2021. Disponível em: <https://arxiv.org/abs/2110.09839>. Acesso em: 4 mar. 2025.

YUCER, Seyma; TEKTAS, Furkan; AL MOUBAYED, Noura; BRECKON, Toby P. Racial bias within face recognition: a survey. **arXiv preprint**, arXiv:2305.00817, 2023. Disponível em: <https://arxiv.org/abs/2305.00817>. Acesso em: 4 mar. 2025.

ZHANG, X.; GAO, Y. Face recognition across pose: a review. **Pattern Recognition**, v. 42, p. 2876–2896, 2009. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0031320309001538>. Acesso em: 19 mar. 2025.

TRANSFORMACIÓN DIGITAL, INTELIGENCIA ARTIFICIAL Y AUTOMATIZACIÓN: RECONFIGURACIÓN DE LOS MODELOS DE GESTIÓN ORGANIZACIONAL

Pilar Nely Flores Salazar
Universidad Nacional del Altiplano Puno

Wilderth Erick Velasquez Medina
Universidad Nacional del Altiplano Puno

Marisol Gonzaga Zirena Cano
Universidad Nacional del Altiplano Puno

Bertelly Turpo Aliaga
Universidad Nacional del Altiplano Puno

Delia Concepcion Cahuana Pacco
Universidad Nacional del Altiplano Puno

RESUMEN

El presente estudio analiza cómo la transformación digital, impulsada por la inteligencia artificial y la automatización, está reconfigurando de manera profunda los modelos de gestión organizacional y fortaleciendo la competitividad empresarial en el contexto del siglo XXI. Se profundiza en la incorporación estratégica de tecnologías emergentes como la inteligencia artificial, la automatización robótica de procesos, el big data, el Internet de las cosas (IoT) y la cadena de bloques (blockchain) en los sistemas organizacionales, evaluando sus efectos en la eficiencia operativa, la capacidad de innovación y la construcción de ventajas competitivas sostenibles. La investigación revela que aquellas organizaciones que adoptan estrategias integrales de transformación digital, alineando de forma coherente procesos, personas y tecnología bajo principios éticos y de sostenibilidad, alcanzan mayores niveles de adaptabilidad, resiliencia y creación de valor. Asimismo, se destacan factores críticos de éxito, entre ellos el liderazgo transformacional, la cultura digital, la gobernanza ética de la inteligencia artificial, la gestión eficaz del cambio organizacional y el desarrollo de competencias digitales clave. Los hallazgos ponen de manifiesto que una transformación digital efectiva requiere enfoques integradores que trasciendan la mera incorporación tecnológica, demandando un rediseño profundo de procesos, una reestructuración organizativa y la consolidación de nuevas capacidades dinámicas.

Palabras clave: transformación digital; inteligencia artificial; gestión organizacional; innovación tecnológica.

INTRODUCCIÓN

La transformación digital representa uno de los fenómenos más disruptivos y determinantes en la evolución de la gestión organizacional contemporánea, configurándose como un imperativo estratégico fundamental para la supervivencia, competitividad y sostenibilidad empresarial en la era de la Cuarta Revolución Industrial (Schwab, 2017). Esta transformación trasciende ampliamente la simple adopción de tecnologías digitales, constituyendo un proceso integral de reconfiguración organizacional que abarca la redefinición de modelos de negocio, la reestructuración de procesos operativos, la transformación de la cultura corporativa, y el desarrollo de nuevas capacidades dinámicas que permitan a las organizaciones navegar efectivamente en entornos caracterizados por la volatilidad, incertidumbre, complejidad y ambigüedad exponencial.

El concepto de transformación digital, según (Vial, 2019), se define como un proceso organizacional que tiene como objetivo mejorar una entidad mediante la activación de cambios significativos en sus propiedades a través de combinaciones de tecnologías de información, computación, comunicación y conectividad. Esta definición encapsula la naturaleza multidimensional del fenómeno, que integra no solo componentes tecnológicos como inteligencia artificial, automatización robótica de procesos, analytics avanzado, Internet de las Cosas, computación en la nube y blockchain, sino también elementos organizacionales críticos como liderazgo, cultura, talento, procesos y gobernanza (Matt *et al.*, 2015).

La inteligencia artificial emerge como el núcleo tecnológico central de esta transformación, proporcionando capacidades sin precedentes para el procesamiento automatizado de información, la generación de insights predictivos, la optimización de procesos complejos y la personalización masiva de productos y servicios (Russell & Norvig, 2020). Las aplicaciones de IA en contextos organizacionales abarcan desde sistemas de recomendación y chatbots inteligentes hasta algoritmos de machine learning para análisis predictivo, procesamiento de lenguaje natural para automatización de documentos, visión computacional para control de calidad, y sistemas de soporte a la decisión basados en analytics avanzado.

La automatización, complementariamente, se configura como el mecanismo operativo que materializa las capacidades de la IA en procesos organizacionales tangibles, eliminando tareas repetitivas, reduciendo errores humanos, acelerando ciclos de producción y liberando capital humano para actividades de mayor valor agregado (Brynjolfsson & McAfee, 2014). La automatización robótica de procesos (RPA) representa una de las manifestaciones más inmediatas y escalables de esta tendencia, permitiendo la automatización de procesos administrativos y operativos sin requerir modificaciones sustanciales en los sistemas de información existentes.

La convergencia de estas tecnologías genera oportunidades exponenciales para la creación de valor organizacional, pero simultáneamente plantea desafíos complejos relacionados con la gestión del cambio, el desarrollo de competencias digitales, la gobernanza ética de la IA, la ciberseguridad, la privacidad de datos y la sostenibilidad social de la automatización (Davenport & Ronanki, 2018). Las organizaciones exitosas en esta transformación son aquellas que logran articular coherentemente estas dimensiones tecnológicas y organizacionales en estrategias integrales que generen valor sostenible para múltiples stakeholders.

La importancia estratégica de la transformación digital se evidencia en su capacidad para generar ventajas competitivas sostenibles a través de múltiples mecanismos: mejora de la eficiencia operativa mediante automatización inteligente, incremento de la calidad y consistencia de productos y servicios, aceleración de procesos de innovación, personalización masiva de experiencias de cliente, optimización de cadenas de suministro, y desarrollo de nuevos modelos de negocio digitales (Porter & Heppelmann, 2014). Adicionalmente, la transformación digital fortalece la resiliencia organizacional al diversificar canales de valor, reducir dependencias operativas y crear capacidades de adaptación rápida ante cambios del entorno.

El objetivo del presente capítulo es analizar comprehensivamente cómo la transformación digital, articulada con la inteligencia artificial y la automatización, reconfigura los modelos de gestión organizacional y potencia la competitividad empresarial, proporcionando un marco conceptual y metodológico para el diseño e implementación de estrategias integrales que articulen procesos, personas y tecnología de forma ética, eficiente y sostenible en el contexto organizacional del siglo XXI.

DESARROLLO

El ecosistema tecnológico que sustenta la transformación digital contemporánea se caracteriza por la convergencia sinérgica de múltiples tecnologías emergentes que, cuando se integran estratégicamente, generan capacidades organizacionales exponencialmente superiores a la suma de sus componentes individuales (Kane *et al.*, 2015). Esta convergencia tecnológica constituye la infraestructura fundamental sobre la cual las organizaciones construyen nuevas capacidades competitivas y redefinen sus proposiciones de valor en mercados digitalmente mediados.

La inteligencia artificial representa el núcleo tecnológico central de esta transformación, abarcando un espectro amplio de tecnologías y aplicaciones que incluyen machine learning, deep learning, procesamiento de lenguaje natural, visión computacional, sistemas expertos, y algoritmos de optimización (LeCun *et al.*, 2015). El machine learning, particularmente, ha emergido como la tecnología más impactante para aplicaciones empresariales, proporcionando capacidades para el reconocimiento de patrones complejos, la predicción de comportamientos futuros, la optimización de procesos en tiempo real, y la automatización de decisiones basadas en análisis de grandes volúmenes de datos estructurados y no estructurados.

La automatización robótica de procesos (RPA) constituye una manifestación práctica inmediatamente implementable de la IA aplicada, permitiendo la automatización de tareas repetitivas, basadas en reglas, que tradicionalmente requerían intervención humana (Willcocks *et al.*, 2015). La RPA se distingue por su capacidad para interactuar con sistemas de información existentes a través de interfaces de usuario, eliminando la necesidad de modificaciones costosas en arquitecturas de software legacy. Esta característica ha facilitado la adopción masiva de RPA en sectores como servicios financieros, seguros, telecomunicaciones y manufactura, donde los procesos administrativos estandarizados representan oportunidades significativas de eficiencia.

La evolución hacia la hiperautomatización representa la siguiente frontera de la automatización organizacional, integrando RPA con tecnologías de IA más sofisticadas como procesamiento de lenguaje natural, visión computacional y analytics predictivo para crear sistemas de automatización inteligente capaces

de manejar procesos complejos, no estructurados y variables (Bharadwaj *et al.*, 2013). Esta evolución permite la automatización de procesos que requieren interpretación de documentos no estructurados, toma de decisiones basada en múltiples variables, y adaptación dinámica a condiciones cambiantes del entorno operativo.

El big data y analytics avanzado proporcionan la infraestructura informacional fundamental para la transformación digital, permitiendo a las organizaciones capturar, almacenar, procesar y analizar volúmenes masivos de datos estructurados y no estructurados para generar insights accionables que informen la toma de decisiones estratégicas y operativas (Chen *et al.*, 2012). La capacidad de procesar y analizar datos en tiempo real ha emergido como una ventaja competitiva crítica, permitiendo respuestas inmediatas a cambios en demanda de mercado, optimización dinámica de operaciones, y personalización en tiempo real de experiencias de cliente.

La computación en la nube constituye la plataforma tecnológica habilitadora que democratiza el acceso a capacidades computacionales avanzadas, proporcionando infraestructura escalable, flexible y económicamente eficiente para la implementación de soluciones de IA y automatización (Mell & Grance, 2011). Los servicios de nube especializados en IA han reducido significativamente las barreras de entrada para la adopción de IA en organizaciones de diferentes tamaños y sectores.

El Internet de las Cosas (IoT) expande exponencialmente la superficie de captura de datos organizacionales, conectando dispositivos físicos, sensores, maquinaria y productos para crear ecosistemas de información en tiempo real que alimentan sistemas de IA y automatización (Atzori *et al.*, 2010). La convergencia de IoT con edge computing permite el procesamiento local de datos, reduciendo latencias y mejorando la responsividad de sistemas automatizados críticos para operaciones en tiempo real.

Blockchain emerge como una tecnología habilitadora para la confianza digital, proporcionando mecanismos descentralizados de verificación, inmutabilidad y trazabilidad que son fundamentales para aplicaciones que requieren transparencia, auditabilidad y eliminación de intermediarios (Tapscott & Tapscott, 2016). En contextos organizacionales, blockchain facilita la automatización

de contratos inteligentes, la verificación automatizada de transacciones, y la creación de ecosistemas de confianza digital entre múltiples participantes.

AIOps (Artificial Intelligence for IT Operations) representa la aplicación específica de IA para la gestión y optimización de infraestructuras de TI, proporcionando capacidades para el monitoreo predictivo, la detección automatizada de anomalías, la resolución proactiva de problemas, y la optimización continua del rendimiento de sistemas tecnológicos (Dang *et al.*, 2019). AIOps se configura como una capacidad crítica para organizaciones que dependen intensivamente de infraestructuras tecnológicas complejas y distribuidas.

La formulación e implementación de estrategias efectivas para la integración tecnológica en procesos de transformación digital constituye uno de los desafíos más complejos y críticos que enfrentan las organizaciones contemporáneas, requiriendo enfoques holísticos que articulen coherentemente dimensiones tecnológicas, organizacionales, culturales y estratégicas en marcos integrales de transformación. El éxito en esta integración depende fundamentalmente de la capacidad organizacional para alinear iniciativas tecnológicas con objetivos estratégicos del negocio, crear sinergias entre diferentes tecnologías, y gestionar efectivamente las interdependencias complejas entre sistemas, procesos y personas.

El alineamiento estratégico entre objetivos de negocio e iniciativas de transformación digital requiere la articulación explícita de cómo las tecnologías emergentes contribuyen a la generación de valor organizacional, el fortalecimiento de ventajas competitivas, y el logro de objetivos estratégicos específicos. Este alineamiento debe manifestarse en la definición clara de casos de uso específicos que demuestren retorno de inversión medible, en la priorización de iniciativas basada en impacto estratégico y factibilidad técnica, y en la integración de métricas de transformación digital en sistemas de medición de desempeño organizacional.

La creación de una visión digital coherente constituye el fundamento conceptual para la integración tecnológica efectiva, proporcionando un marco orientador que articule aspiraciones organizacionales con capacidades tecnológicas específicas. Esta visión debe especificar cómo la organización utilizará tecnologías digitales para crear valor diferencial, mejorar experiencias de cliente, optimizar operaciones internas, y desarrollar nuevas capacidades

competitivas. La visión digital efectiva trasciende declaraciones aspiracionales para incluir roadmaps específicos, métricas de progreso, y marcos de gobernanza para la toma de decisiones de inversión tecnológica.

Los roles de liderazgo digital emergentes, particularmente Chief Digital Officer (CDO) y Chief AI Officer (CAIO), se configuran como catalizadores organizacionales críticos para la integración tecnológica exitosa, proporcionando liderazgo especializado, coordinación interfuncional, y evangelización de la transformación digital a través de la organización (Singh & Hess, 2017). El CDO típicamente asume responsabilidades para la definición de estrategia digital, la coordinación de iniciativas de transformación, el desarrollo de capacidades digitales, y la gestión de alianzas tecnológicas estratégicas. El CAIO se especializa en la governance de IA, la ética algorítmica, la gestión de riesgos de IA, y la maximización del valor organizacional de inversiones en inteligencia artificial.

La arquitectura empresarial digital constituye el framework técnico que habilita la integración coherente de múltiples tecnologías en ecosistemas organizacionales complejos, proporcionando principios, estándares y patrones para la interoperabilidad, escalabilidad y mantenibilidad de soluciones tecnológicas (Ross *et al.*, 2006). Una arquitectura empresarial efectiva debe facilitar la integración fluida entre sistemas legacy y tecnologías emergentes, habilitar la reutilización de componentes tecnológicos a través de diferentes aplicaciones, y proporcionar flexibilidad para la incorporación futura de nuevas tecnologías sin disrupciones significativas en operaciones existentes.

El desarrollo de capacidades organizacionales específicas para la transformación digital requiere inversiones sistemáticas en competencias técnicas, habilidades analíticas, y capacidades de gestión del cambio que permitan a la organización maximizar el valor de inversiones tecnológicas (Teece *et al.*, 1997). Estas capacidades incluyen competencias en ciencia de datos para el aprovechamiento efectivo de analytics avanzado, habilidades de diseño de experiencia de usuario para la creación de interfaces digitales intuitivas, competencias en gestión de proyectos ágiles para la implementación iterativa de soluciones tecnológicas, y capacidades de gestión del cambio para facilitar la adopción organizacional de nuevas tecnologías.

El diseño e implementación de una hoja de ruta (roadmap) estratégica para la transformación digital constituye un proceso metodológico crítico que

determina fundamentalmente el éxito de iniciativas de integración tecnológica organizacional, requiriendo enfoques sistemáticos que equilibren aspiraciones estratégicas con realidades operativas, limitaciones de recursos y capacidades organizacionales existentes. Esta hoja de ruta debe articular coherentemente la secuenciación temporal de iniciativas, la asignación óptima de recursos, la gestión de interdependencias tecnológicas, y la mitigación de riesgos asociados con la implementación de tecnologías emergentes en contextos organizacionales complejos.

La identificación y priorización de procesos organizacionales para la integración de IA y automatización requiere análisis sistemáticos que evalúen múltiples dimensiones: potencial de impacto en eficiencia operativa, complejidad técnica de implementación, disponibilidad de datos de calidad, madurez de tecnologías aplicables, y alineamiento con objetivos estratégicos organizacionales. Los procesos más apropiados para automatización inicial típicamente se caracterizan por alta frecuencia de ejecución, reglas de decisión claras y estables, bajo nivel de excepción y variabilidad, y interfaces bien definidas con sistemas de información existentes.

La metodología de evaluación de procesos debe incorporar frameworks estructurados como la matriz de impacto versus factibilidad, que posiciona iniciativas potenciales en cuadrantes que facilitan la priorización basada en criterios objetivos. Los procesos de alto impacto y alta factibilidad constituyen candidatos ideales para implementación en fases iniciales, proporcionando victorias tempranas que generen momentum organizacional y demuestren valor tangible de inversiones en transformación digital. Los procesos de alto impacto pero baja factibilidad requieren estrategias de desarrollo de capacidades y pueden considerarse para fases posteriores del roadmap.

El diseño de fases de implementación debe seguir principios de escalamiento progresivo que inicien con pilotos controlados en áreas de bajo riesgo, avancen hacia implementaciones departamentales, y culminen en despliegues organizacionales amplios que maximicen sinergias entre diferentes iniciativas de transformación. Cada fase debe incluir objetivos específicos, métricas de éxito claramente definidas, criterios de decisión para avanzar a la siguiente fase, y mecanismos de rollback en caso de identificar riesgos o problemas significativos durante la implementación.

La fase de pilotaje debe enfocarse en la validación de hipótesis técnicas y de negocio en contextos controlados que permitan aprendizaje rápido y ajustes iterativos sin impactos significativos en operaciones críticas. Los pilotos efectivos típicamente incluyen casos de uso específicos con métricas de éxito predefinidas, equipos multidisciplinarios con representación de TI y negocio, y duraciones limitadas que faciliten evaluación rápida de resultados. La documentación sistemática de lecciones aprendidas durante pilotos proporciona insumos valiosos para el refinamiento de enfoques de implementación en fases subsecuentes.

Los ejes transversales de gobernanza, cultura, talento y datos constituyen dimensiones críticas que deben desarrollarse sistemáticamente a lo largo de todas las fases del roadmap, proporcionando la infraestructura organizacional fundamental para el éxito sostenible de la transformación digital. La gobernanza debe establecer marcos para la toma de decisiones coordinada, la gestión de riesgos, y la optimización de inversiones. La cultura debe evolucionar hacia orientaciones digitales que valoren la experimentación, el aprendizaje continuo, y la colaboración interfuncional. El talento debe desarrollar competencias digitales específicas y capacidades de adaptación a tecnologías emergentes. Los datos deben gestionarse como activos estratégicos con arquitecturas, procesos y governance que faciliten su aprovechamiento efectivo para analytics e IA.

La gestión efectiva del cambio organizacional constituye un factor determinante crítico para el éxito de iniciativas de transformación digital, requiriendo enfoques multidimensionales que aborden simultáneamente aspectos culturales, estructurales, procesales y tecnológicos de la transformación, mientras gestionan proactivamente resistencias naturales al cambio y facilitan la adopción organizacional de nuevas tecnologías, procesos y formas de trabajo (Kotter, 2012). La complejidad inherente de la transformación digital amplifica significativamente los desafíos tradicionales de gestión del cambio, requiriendo metodologías especializadas que reconozcan la naturaleza iterativa, experimental y evolutiva de los procesos de digitalización organizacional.

El desarrollo de una cultura digital organizacional representa uno de los elementos más críticos y complejos de la transformación, requiriendo cambios fundamentales en mindsets, comportamientos, valores y prácticas organizacionales hacia orientaciones que privilegien la experimentación, el aprendizaje continuo, la colaboración interfuncional, la orientación al cliente,

y la adaptabilidad ante cambios tecnológicos acelerados. Esta transformación cultural debe manifestarse en la valoración del fracaso como oportunidad de aprendizaje, la adopción de metodologías ágiles para el desarrollo de productos y servicios, la promoción de la toma de decisiones basada en datos, y el fomento de la innovación y creatividad en todos los niveles organizacionales.

Las competencias digitales emergentes requieren programas sistemáticos de desarrollo de talento que aborden múltiples niveles de sofisticación: alfabetización digital básica para todos los empleados, competencias analíticas intermedias para roles que requieren interpretación de datos, y expertise técnico avanzado para especialistas en IA, ciencia de datos, automatización y ciberseguridad. Estos programas deben incluir componentes de formación formal, aprendizaje experiencial a través de proyectos reales, mentorías con expertos internos y externos, y certificaciones en tecnologías específicas que validen competencias adquiridas.

La gobernanza ética de la inteligencia artificial emerge como una dimensión crítica que requiere frameworks organizacionales específicos para asegurar que las implementaciones de IA respeten principios éticos fundamentales, mitiguen riesgos de sesgo algorítmico, protejan privacidad y derechos de datos, y mantengan transparencia en procesos de toma de decisiones automatizados (Russell, 2019). Esta gobernanza debe incluir comités éticos multidisciplinarios, principios éticos explícitos para el desarrollo y implementación de IA, procesos de auditoría algorítmica para detectar sesgos y discriminación, y mecanismos de accountability para decisiones automatizadas que afecten individuos o grupos.

Los frameworks de IA responsable deben abordar múltiples dimensiones de riesgo: fairness para asegurar tratamiento equitativo de diferentes grupos demográficos, explainability para proporcionar transparencia en decisiones algorítmicas, robustness para garantizar funcionamiento confiable en condiciones variables, privacy para proteger información personal sensible, y human agency para mantener control humano sobre decisiones críticas (Floridi *et al.*, 2018). La implementación efectiva de estos frameworks requiere colaboración estrecha entre equipos técnicos, legales, éticos y de negocio para asegurar que consideraciones éticas se integren sistemáticamente en ciclos de desarrollo de soluciones de IA.

La operacionalización exitosa de pilotos hacia operaciones continuas requiere capacidades organizacionales específicas para la gestión de tecnologías en producción, incluyendo monitoreo de performance, gestión de incidentes, actualizaciones y mantenimiento de sistemas, y optimización continua basada en datos de uso real. Esta operacionalización debe incluir definición de Service Level Agreements específicos, procesos de escalamiento para resolución de problemas, y capacidades de rollback para mitigar impactos de actualizaciones problemáticas.

AIOps (Artificial Intelligence for IT Operations) se configura como una capacidad crítica para la gestión escalable de infraestructuras tecnológicas complejas, proporcionando automatización inteligente para monitoreo, detección de anomalías, diagnóstico de problemas, y optimización de performance de sistemas distribuidos. La implementación de AIOps facilita la transición de modelos reactivos de gestión de TI hacia enfoques predictivos y proactivos que anticipen problemas potenciales y optimicen automáticamente configuraciones para mantener niveles óptimos de servicio.

Medición de Impacto y Estudios de Caso

La medición sistemática del impacto de iniciativas de transformación digital constituye un imperativo crítico para la validación de inversiones tecnológicas, la optimización continua de implementaciones, y la demostración de valor organizacional ante stakeholders internos y externos. Los sistemas de medición efectivos deben incorporar métricas financieras tradicionales como retorno de inversión y reducción de costos, complementadas con indicadores específicos de transformación digital que capturen mejoras en eficiencia operativa, calidad de experiencias de cliente, velocidad de innovación, y capacidades de adaptación organizacional.

Los indicadores clave de performance (KPIs) para transformación digital deben estructurarse en múltiples dimensiones que reflejen la naturaleza multifacética del impacto digital: métricas financieras (ROI, reducción de costos, incremento de ingresos), métricas operacionales (eficiencia de procesos, reducción de tiempos de ciclo, calidad de productos), métricas de cliente (satisfacción, Net Promoter Score, retención), métricas de empleados (engagement, productividad, satisfacción), y métricas de innovación (tiempo de lanzamiento de productos, número de iniciativas digitales, adopción de tecnologías emergentes).

Los estudios de caso de implementaciones exitosas proporcionan evidencia empírica valiosa sobre factores críticos de éxito, mejores prácticas, y enfoques efectivos para diferentes contextos organizacionales e industriales. El caso de General Electric ilustra cómo una organización industrial tradicional logró transformarse en una empresa digital a través de inversiones masivas en IoT, analytics avanzado, y plataformas digitales, creando nuevos modelos de negocio basados en servicios digitales que complementan productos físicos tradicionales.

Amazon representa un paradigma de transformación digital nativa que ha revolucionado múltiples industrias a través de la aplicación sistemática de tecnologías emergentes, incluyendo IA para sistemas de recomendación, automatización para logística y fulfillment, y cloud computing para escalabilidad operacional. Las capacidades digitales de Amazon han facilitado diversificación hacia sectores completamente nuevos, desde cloud computing hasta inteligencia artificial conversacional, demostrando el potencial de tecnologías digitales para crear nuevas fuentes de ventaja competitiva.

Netflix ejemplifica la transformación de modelos de negocio tradicionales hacia plataformas digitales que aprovechan datos para personalización masiva, utilizando algoritmos de machine learning para optimizar recomendaciones de contenido, decisiones de producción, y estrategias de adquisición de contenido basadas en patrones de comportamiento de usuarios. La capacidad de Netflix para procesar y analizar datos de viewing en tiempo real ha creado ventajas competitivas sostenibles en creación y distribución de contenido digital.

Los aprendizajes de implementaciones fallidas proporcionan insights igualmente valiosos sobre riesgos comunes y factores que contribuyen al fracaso de iniciativas de transformación digital. Los errores más frecuentes incluyen insuficiente alineamiento entre estrategia digital y objetivos de negocio, subestimación de complejidades de gestión del cambio, inadecuada inversión en desarrollo de competencias digitales, y ausencia de governance efectiva para coordinación de iniciativas múltiples.

REFLEXIONES, RECOMENDACIONES Y CONCLUSIONES

Los hallazgos de esta investigación confirman de manera contundente que la transformación digital, articulada estratégicamente con inteligencia artificial

y automatización, representa no simplemente una modernización tecnológica opcional, sino un imperativo existencial fundamental para la supervivencia, competitividad y sostenibilidad organizacional en el contexto de la economía digital del siglo XXI. La evidencia analizada demuestra consistentemente que las organizaciones que logran implementar exitosamente estrategias integrales de transformación digital, articulando coherentemente tecnología, procesos, personas y cultura, alcanzan niveles superiores de eficiencia operativa, innovación, adaptabilidad y generación de valor sostenible comparadas con organizaciones que mantienen enfoques tradicionales de gestión.

La transformación digital exitosa trasciende ampliamente la simple adopción de tecnologías emergentes, requiriendo reconfiguración fundamental de modelos organizacionales, rediseño de procesos operativos, desarrollo de nuevas competencias y capacidades, y evolución de culturas corporativas hacia orientaciones digitales que privilegien la experimentación, el aprendizaje continuo, y la adaptación constante ante cambios tecnológicos acelerados. Esta naturaleza multidimensional de la transformación digital demanda enfoques holísticos e integrales que aborden simultáneamente aspectos tecnológicos, organizacionales, culturales y estratégicos en marcos coherentes de cambio organizacional.

El alineamiento estratégico entre objetivos organizacionales e iniciativas de transformación digital emerge como un factor crítico determinante del éxito, requiriendo articulación explícita de cómo tecnologías específicas contribuyen a la generación de valor organizacional, el fortalecimiento de ventajas competitivas, y el logro de objetivos estratégicos específicos. Las organizaciones más exitosas en transformación digital son aquellas que logran traducir aspiraciones digitales en roadmaps específicos con casos de uso claramente definidos, métricas de impacto medibles, y marcos de governance que faciliten la coordinación efectiva de iniciativas múltiples y complejas.

La gestión efectiva del cambio organizacional se configura como una capacidad fundamental para la transformación digital, requiriendo metodologías especializadas que reconozcan la naturaleza iterativa, experimental y evolutiva de los procesos de digitalización. El desarrollo de culturas digitales que valoren la experimentación, el aprendizaje de errores, la colaboración interfuncional, y la adaptabilidad ante incertidumbre representa uno de los desafíos más complejos pero críticos de la transformación, requiriendo liderazgo transformacional

consistente y programas sistemáticos de desarrollo de competencias digitales en todos los niveles organizacionales.

La gobernanza ética de la inteligencia artificial emerge como una dimensión crítica que requiere frameworks organizacionales específicos para asegurar que implementaciones de IA respeten principios éticos fundamentales, mitiguen riesgos de sesgo y discriminación, y mantengan transparencia en procesos automatizados de toma de decisiones. La implementación de IA responsable requiere colaboración multidisciplinaria entre equipos técnicos, legales, éticos y de negocio para integrar consideraciones éticas sistemáticamente en ciclos de desarrollo de soluciones tecnológicas.

Las perspectivas futuras de la transformación digital sugieren una evolución hacia paradigmas de inteligencia aumentada donde capacidades humanas y artificiales se complementen sinérgicamente, hacia operaciones autónomas que requieran mínima intervención humana para procesos rutinarios, y hacia ecosistemas digitales interconectados que faciliten colaboración e innovación entre múltiples organizaciones. La inteligencia artificial generativa emerge como una frontera tecnológica particularmente prometedora para la automatización de actividades creativas y analíticas que tradicionalmente requerían expertise humano especializado.

Finalmente, el llamado a la acción para administradores del siglo XXI implica el desarrollo de mindsets de aprendizaje continuo y adaptación constante ante cambios tecnológicos acelerados, la cultivación de competencias digitales que combinen expertise técnico con habilidades de liderazgo y gestión del cambio, y el compromiso con principios éticos que aseguren que la transformación digital contribuya positivamente al bienestar social y la sostenibilidad ambiental. Los líderes organizacionales del futuro deben configurarse como arquitectos de transformación que diseñen y construyan organizaciones adaptativas, resilientes e innovadoras capaces de prosperar en entornos de cambio permanente y complejidad creciente.

REFERENCIAS

- Atzori, L., Iera, A., & Morabito, G. (2010). **The internet of things**: A survey. *Computer Networks*, 54(15), 2787–2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly*, 37(2), 471–482. <https://doi.org/10.25300/MISQ/2013/37:2.3>
- Brynjolfsson, E., & McAfee, A. (2014). **The second machine age**: Work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company. <https://doi.org/10.12691/ajie-2-4-2>
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188. <https://doi.org/10.2307/41703503>
- Dang, Y., Lin, Q., & Huang, P. (2019). AIOps: Real-world challenges and research innovations. Proceedings of the 41st **International Conference on Software Engineering**, 4–5. <https://doi.org/10.1109/ICSE-Companion.2019.00023>
- Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116. <https://doi.org/10.1287/mnsc.2018.3204>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Kane, G. C., Phillips, A. N., Copulsky, J., & Andrus, G. (2015). The digital matrix: New rules for business transformation through technology. *MIT Press*. https://doi.org/10.1162/DAED_a_00359
- Kotter, J. P. (2012). Leading change. Harvard Business Review Press. <https://doi.org/10.1002/9781119204183.ch1>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Matt, C., Hess, T., & Benlian, A. (2015). Digital transformation strategies. *Business & Information Systems Engineering*, 57(5), 339–343. <https://doi.org/10.1007/s12599-015-0401-5>
- Mell, P., & Grance, T. (2011). **The NIST definition of cloud computing**. <https://doi.org/10.6028/NIST.SP.800-145>
- Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), 64–88. <https://doi.org/10.1287/orsc.2015.1020>
- Ross, J. W., Weill, P., & Robertson, D. (2006). Enterprise architecture as strategy: Creating a foundation for business execution. *Harvard Business Review Press*. <https://doi.org/10.5465/amr.2007.26586092>
- Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. *Viking Press*. <https://doi.org/10.1038/s41586-019-1724-z>
- Russell, S., & Norvig, P. (2020). **Artificial intelligence**: A modern approach (4th ed.). Pearson. <https://doi.org/10.1017/S0269888921000066>
- Schwab, K. (2017). The fourth industrial revolution. *Currency Books*. <https://doi.org/10.1108/dprg-12-2017-0062>
- Singh, A., & Hess, T. (2017). How chief digital officers promote the digital transformation of their companies. *MIS Quarterly Executive*, 16(1), 1–17. <https://doi.org/10.4018/IJISMD.2017040101>

Tapscott, D., & Tapscott, A. (2016). Blockchain revolution: How the technology behind Bitcoin is changing money, business, and the world. **Penguin Random House**. <https://doi.org/10.1108/SL-05-2017-0047>

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. **Strategic Management Journal**, 18(7), 509–533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. **Journal of Strategic Information Systems**, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>

Willcocks, L. P., Lacity, M., & Craig, A. (2015). The IT function and robotic process automation. **LSE Business Review**. <https://doi.org/10.1108/JSIT-09-2016-0061>

SOBRE O ORGANIZADOR

Ernane Rosa Martins

Pós-Doutorado em E-learning pela Universidade Fernando Pessoa (UFP). Doutor em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa (UFP), em Porto/Portugal, reconhecido como equivalente ao curso de Doutorado em Ciência da Informação, da UnB. Mestre em Engenharia de Produção e Sistemas pela UCG, possui Pós-Graduação em Tecnologia em Gestão da Informação, Graduação em Ciência da Computação e Graduação em Sistemas de Informação. Professor de Informática, com dedicação exclusiva, do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia), ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador líder do grupo de pesquisa NITE (Núcleo de Inovação, Tecnologia e Educação), certificado pelo IFG no CNPq. Membro do Conselho Editorial da Editora Científica Digital.

Lattes: <http://lattes.cnpq.br/5566965064833628>

ÍNDICE REMISSIVO

Símbolos

338/23: 45, 46, 48, 49, 50, 51, 55, 58, 59

A

Algoritmos: 10, 20, 21, 22, 23, 24, 25, 27, 32, 33, 39, 44, 50, 60, 64, 75, 80, 81, 87, 88, 89, 90, 92, 93, 96, 97, 98, 99, 100, 106, 107, 109, 110, 111, 112, 114, 115, 119, 123, 126, 127, 129, 130, 131, 132, 133, 139, 145, 146, 147, 149, 150, 151, 152, 153, 154, 155, 157, 158, 165, 167, 175

Análisis Financiero: 62, 63, 64, 70

Aplicação: 10, 12, 17, 21, 25, 38, 48, 49, 52, 60, 86, 88, 89, 91, 92, 96, 97, 99, 100, 107, 108, 109, 111, 112, 119, 120, 126, 128, 130, 133, 134, 135, 136, 137, 139, 140, 147

Aprendizado de Máquina: 87, 94, 101, 106, 107, 110, 122, 143, 157

Aprendizado por Reforço: 8, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 29, 93

Aquicultura: 92, 93, 105, 106, 107, 108, 109, 110, 111, 113, 115, 116, 117, 120, 123

Artificial: 10, 11, 12, 14, 15, 17, 18, 27, 28, 29, 30, 31, 32, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 57, 58, 59, 60, 61, 63, 71, 79, 81, 82, 85, 86, 99, 103, 105, 106, 107, 108, 109, 111, 113, 115, 116, 119, 120, 121, 123, 126, 127, 130, 131, 132, 133, 137, 138, 143, 144, 146, 157, 162, 163, 164, 165, 166, 167, 169, 170, 173, 174, 175, 177, 178

Automação: 38, 89, 91, 99, 100, 101, 102, 106, 107, 108, 111, 112, 113, 131

C

Comportamento Adaptativo: 8, 9, 10, 11, 14, 21

Confiança: 30, 31, 32, 33, 34, 37, 40, 41, 42, 58, 59, 96, 113, 119, 120, 131, 133

Contabilidade Gerencial: 62, 63, 64, 65, 67, 73, 79, 80, 81, 84

E

Ética: 30, 31, 32, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 56, 58, 59, 60, 61, 89, 96, 97, 98, 99, 101, 102, 107, 111, 117, 119, 123, 126, 133, 136, 137, 139, 140, 164, 166, 170, 173, 177

Ética: 30, 31, 32, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 56, 58, 59, 60, 61, 89, 96,

97, 98, 99, 101, 102, 107, 111, 117, 119, 123, 126, 133, 136, 137, 139, 140, 164, 166, 170, 173, 177

Explicabilidade: 31, 32, 33, 34, 37, 39, 40, 41, 43, 45, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 98

G

Gestão de Riesgos: 63, 170, 172

Gestão Organizacional: 163, 164, 165, 166

I

IA: 9, 10, 12, 15, 16, 27, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50, 54, 55, 56, 58, 59, 60, 61, 62, 71, 83, 86, 89, 90, 91, 92, 93, 94, 96, 97, 98, 99, 100, 101, 102, 103, 106, 107, 108, 109, 110, 111, 112, 113, 114, 117, 119, 121, 122, 123, 130, 132, 165, 166, 167, 168, 169, 170, 171, 172, 173, 175, 177

Innovación Tecnológica: 164

Inteligência: 10, 11, 12, 14, 15, 17, 18, 27, 29, 30, 31, 32, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 57, 58, 59, 60, 61, 85, 86, 89, 99, 103, 105, 106, 107, 108, 109, 111, 113, 115, 116, 119, 120, 121, 123, 126, 127, 130, 137, 143, 146, 148, 157, 162

Inteligencia Artificial: 63, 71, 79, 81, 163, 164, 165, 166, 167, 170, 173, 175, 177

Inteligência Artificial: 10, 11, 12, 14, 15, 17, 18, 27, 29, 30, 31, 32, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 57, 58, 59, 60, 61, 85, 86, 99, 105, 106, 107, 108, 109, 111, 113, 115, 116, 119, 120, 121, 123, 126, 127, 130, 137, 143, 146, 157, 162

Inteligência Artificial: 10, 11, 12, 14, 15, 17, 18, 27, 29, 30, 31, 32, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 57, 58, 59, 60, 61, 85, 86, 99, 105, 106, 107, 108, 109, 111, 113, 115, 116, 119, 120, 121, 123, 126, 127, 130, 137, 143, 146, 157, 162

J

Jogos Roguelike/Roguelite: 9, 10, 11, 12, 21

N

NPCs Adaptativos: 9, 13, 14, 21, 25, 26, 27

O

Oceanografia: 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 98, 99, 100, 101, 102, 103

P

Projeto-lei nº 2: 46

R

Racismo Algorítmico: 125, 126, 127, 130, 132, 139, 143, 154, 162

Reconhecimento Facial: 125, 126, 127, 128, 129,

ÍNDICE REMISSIVO

130, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162

S

Sector Pesquero: 62, 63, 64, 67, 68, 69, 70, 80, 81, 83

Segurança Pública: 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 137, 138, 139, 140, 142, 147, 148, 149, 158, 159, 160

Sustentabilidade: 48, 99, 102, 103, 106, 107, 109, 110, 111, 113, 114, 115, 116, 119, 120, 122, 123

T

Transformación Digital: 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 174, 175, 176, 177

V

Variáveis Demográficas: 146, 147, 148, 152, 156, 158

Vieses: 31, 32, 33, 34, 37, 39, 40, 43, 47, 54, 57, 58, 97, 100, 126, 127, 129, 130, 131, 132, 135, 136, 137, 138, 139, 140, 147, 148, 149, 153, 154, 155, 156, 157, 158



científica digital



VENDA PROIBIDA - ACESSO LIVRE - OPEN ACCESS

