UNIVERSIDADE FEDERAL DO SUL DA BAHIA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA BAHIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS E TECNOLOGIAS AMBIENTAIS

ÊMERIS SILVA SANTOS

ODSBAHIA-PTBR: UM MODELO DE PROCESSAMENTO DE LINGUAGEM NATURAL PARA APOIAR OS 20 OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL

PORTO SEGURO
2025

ÊMERIS SILVA SANTOS

ODSBAHIA-PTBR: UM MODELO DE PROCESSAMENTO DE LINGUAGEM NATURAL PARA APOIAR OS 20 OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL

Dissertação submetida ao Programa de Pós-graduação em Ciências e Tecnologias Ambientais da Universidade Federal do Sul da Bahia e do Instituto Federal de Educação, Ciência e Tecnologia da Bahia, com vistas à obtenção do título de mestre(a) em Ciências e Tecnologias Ambientais.

Orientador(a): Prof. Dr. Leonardo Evangelista Moraes

PORTO SEGURO
2025

Catalogação na Publicação (CIP)
Universidade Federal do Sul da Bahia (UFSB)
Sistema de Bibliotecas (SIBI)

S237o Santos, Êmeris Silva, 1987 –
ODSBAHIA-PTBR: um modelo de processamento de linguagem natural para apoiar os 20 objetivos de desenvolvimento sustentável. / Êmeris Silva Santos. – Porto Seguro, 2025.
72 f.

Orientador: Prof. Dr. Leonardo Evangelista Moraes
Dissertação (Mestrado) – Universidade Federal do Sul da Bahia. Centro de Formação em Ciências Ambientais. Programa de Pós-Graduação em Ciências e Tecnologias Ambientais. Campus Sosígenes Costa.

1. Agenda 2030. 2. Sustentabilidade. 3. Inteligência Artificial. 4. BERT. I. Moraes, Leonardo Evangelista. II. Título.

CDD – 006.35

Elaborado por Lucas Sousa Carvalho - CRB-5/1883

ATA DE DEFESA PÚBLICA DE DISSERTAÇÃO DE MESTRADO NO PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM CIÊNCIAS E TECNOLOGIAS AMBIENTAIS

Aos vinte e quatro dias do mês de fevereiro do ano de 2025, às 08h40min, via webconferência, por meio da sala virtual com link de transmissão https://meet.google.com/https://meet.google.com/zbb-fhpa-efx, reuniram-se os/as membros/as da banca examinadora composta pelos/as docentes Dr. Leonardo Evangelista Moraes (orientador e presidente da banca), Dr. Bilzã Marques de Araujo (membro externo ao Programa), Dra. Guineverre Alvarez Machado de Melo Gomes (membra externa ao Programa), Dra. Maria do Carmo Rebouças da Cruz Ferreira dos Santos (membra externa ao Programa) e Dr. Thiago Gehre Galvão (membro externo ao Programa), a fim de arguirem o/a mestrando/a Êmeris Silva Santos na defesa de sua dissertação de mestrado, cujo trabalho de pesquisa intitula-se "ODSBAHIAPTBR: Um modelo de processamento de linguagem natural para apoiar os 20 objetivos de desenvolvimento sustentável". Aberta a sessão pelo/a presidente da mesma, coube ao candidato, na forma regimental, expor o tema de sua dissertação, dentro do tempo regulamentar, sendo em seguida questionado pelos/as membros/as da banca examinadora, tendo dado as explicações que foram necessárias. Após reunião em sessão secreta, os/as membros/as da banca consideraram a dissertação:

(X) Aprovado com Louvor ( ) Aprovado ( ) Reprovado

BANCA EXAMINADORA:

gov.br Documento assinado digitalmente
LEONARDO EVANGELISTA MORAES
Data: 25/02/2025 15:10:38-0300
Verifique em https://validar.iti.gov.br

Dr. Leonardo Evangelista Moraes (UFSB/PPGCTA)
Orientador/a e Presidente/a da banca

gov.br Documento assinado digitalmente
BILZA MARQUES DE ARAUJO
Data: 28/02/2025 10:28:41-0300
Verifique em https://validar.iti.gov.br

Dr. Bilzã Marques de Araujo (UFSB)
Membro externo

gov.br Documento assinado digitalmente
GUINEVERRE ALVAREZ MACHADO DE MELO GOM
Data: 06/03/2025 13:10:37-0300
Verifique em https://validar.iti.gov.br

Dra. Guineverre Alvarez Machado de Melo Gomes (UFSB)
Membro externo

gov.br Documento assinado digitalmente
MARIA DO CARMO REBOUCAS DA CRUZ FERREIR
Data: 05/03/2025 23:03:14-0300
Verifique em https://validar.iti.gov.br

Dra. Maria do Carmo Rebouças da Cruz Ferreira dos Santos (UFSB)
Membra externa

gov.br Documento assinado digitalmente
THIAGO GEHRE GALVAO
Data: 11/03/2025 17:50:48-0300
Verifique em https://validar.iti.gov.br

Dr. Thiago Gehre Galvão (UnB)
Membro externo

gov.br Documento assinado digitalmente
EMERIS SILVA SANTOS
Data: 13/03/2025 10:13:14-0300
Verifique em https://validar.iti.gov.br

Emeris Silva Santos
Nome do discente

Webconferência, 24 de fevereiro de 2026

PARECERES DA BANCA EXAMINADORA DE DEFESA:

(exceto presidente/a)

MEMBRO/A	SIGLA DA INSTITUIÇÃO	NOTA (0,0–10,0)	PARECER
Dr. Bilzã Marques de Araujo	UFSB	10,0	Aprovado
Dra. Guineverre Alvarez Machado de Melo Gomes	UFSB	10,0	Aprovado
Dra. Maria do Carmo Rebouças da C. Ferreira dos Santos	UFSB	10,0	Aprovado
Dr. Thiago Gehre Galvão	UnB	9,5	Aprovado
CONCEITO FINAL		**9,88**	**Aprovado**

Conforme o Regimento Interno do PPGCTA (2021, p. 26):

> **Art. 86.** Cada membro(a) da banca, exceto a(o) presidenta(e), emitirá parecer por escrito e atribuirá nota entre 0,0 (zero) a 10,0 (dez), correspondente à apresentação e ao trabalho escrito da dissertação.
> § 1º. O conceito final resultará da média aritmética dos valores numéricos concedidos pelos membros da banca examinadora, sendo considerado aprovada(o) a(o) discente que obtiver média final igual ou superior a 7,0 (sete).
> (...)

Em caso de **reprovação**, a banca deve indicar neste espaço as razões que embasaram tal decisão:

Outras **recomendações** da banca examinadora (OPCIONAL):

Embora o trabalho tenha sido aprovado, a banca indicou que a opção de estruturar o documento em capítulos, especialmente com a inclusão de um artigo já publicado, não favorecer a exposição dos argumentos de modo orgânico. Neste sentido, visando a apresentação do documento para o depósito na biblioteca, a banca sugeriu que fosse adotada a estrutura em forma de dissertação com texto corrido, destacando uma construção que evidencie os pontos fortes de argumentação da dissertação, a saber: 1) as universidades e os desafios com as agendas globais de desenvolvimento e 2) a incorporação de ferramentas, a exemplo da inteligência artificial para responder a este desafio.

gov.br
Documento assinado digitalmente
LEONARDO EVANGELISTA MORAES
Data: 13/03/2025 11:10:16-0300
Verifique em https://validar.iti.gov.br

Dr. Leonardo Evangelista Moraes (UFSB/ PPGCTA)
Presidente/a da banca

DEDICATÓRIA

Dedico este trabalho ao meu saudoso pai, Eufrânio, que sempre dizia: 'Meu sonho é ver um filho de pipoqueiro formado em uma faculdade!'. Pai, este é mais um presente para o senhor!

AGRADECIMENTOS

A Deus, autor da vida, por me conceder saúde, graça e determinação para perseverar durante a realização deste trabalho.

À minha mãe, minha esposa e meus filhos, pelo incentivo nos momentos difíceis e pela compreensão diante da minha ausência enquanto me dedicava a esta pesquisa.

Ao professor Dr. Leonardo Moraes, por sua orientação dedicada e pela amizade demonstrada ao longo deste percurso.

Aos professores, por todos os conselhos, pela ajuda e pela paciência com a qual guiaram o meu aprendizado.

A todos que participaram, direta ou indiretamente do desenvolvimento deste trabalho de pesquisa, enriquecendo o meu processo de aprendizado.

SANTOS, Êmeris Silva. **ODSBAHIA-PTBR**: Um modelo de processamento de linguagem natural para apoiar os 20 objetivos de desenvolvimento sustentável. Orientador: Leonardo Evangelista Moraes. 2025. Dissertação (Mestrado em Ciências e Tecnologias Ambientais) – Universidade Federal do Sul da Bahia; Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Porto Seguro, 2025.

RESUMO

Falta menos de uma década para o prazo limite de cumprimento dos Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pela Organização das Nações Unidas (ONU) em 2015 e ainda existem lacunas na disponibilidade de dados e conhecimento para monitorar o progresso em direção aos ODS e avaliar o impacto das iniciativas implementadas. Este estudo investiga como universidades baianas contribuem para a Agenda 2030, utilizando técnicas de Processamento de Linguagem Natural (PLN). O primeiro capítulo apresenta uma fundamentação teórica que explora o conceito de desenvolvimento sustentável e o histórico das principais conferências internacionais que culminaram na formulação da Agenda 2030. Além disso, o capítulo aborda os desafios globais e locais associados à implementação dos ODS, destacando a necessidade de integrar a perspectiva territorial e os conhecimentos tradicionais. Nesse contexto, são introduzidos os conceitos fundamentais de Inteligência Artificial e sua aplicação em iniciativas relacionadas aos ODS. Essa abordagem estabelece uma base sólida para o desenvolvimento das contribuições técnicas e analíticas apresentadas nos capítulos subsequentes, conectando a Agenda 2030 às ferramentas tecnológicas utilizadas neste trabalho. No segundo capítulo, é apresentado um artigo que desenvolve o ODSBahia-PTBR, um modelo de PLN com aprendizado profundo, baseado no BERTimbau, para classificar textos segundo os 17 ODS da ONU, e os ODS 18 (Igualdade Racial), ODS 19 (Arte, Cultura e Comunicação) e ODS 20 (Direito dos Povos Originários e Comunidades Tradicionais), propostos no 'Guia Agenda 2030: integrando ODS, Educação e Sociedade', elaborado em 2020 em parceria entre a UnB e a UNESP. O ODSBahia-PTBR alcançou a precisão de 82%, recall de 72% e F1-score de 77%, além disso, acertou 95% dos rótulos do conjunto de dados OSDG traduzido para o português, demonstrando sua eficácia na tarefa de classificação. No capítulo 3, a pesquisa descritiva e exploratória analisa dados de projetos de pesquisa e extensão da Universidades Federal do Sul da Bahia (UFSB), da Universidade Federal do Recôncavo da Bahia (UFRB) e da Universidade Federal do Oeste da Bahia (UFOB). A análise envolve a aplicação do ODSBahia-PTBR para classificar os projetos quanto aos ODS. Esta pesquisa se justifica pela importância de compreender, avaliar e direcionar os esforços de pesquisa que são desenvolvidos pelas instituições de ensino superior e suas contribuições para a consecução dos ODS em seus territórios de atuação. Este trabalho alinha-se ao PPGCTA, sobretudo à linha de pesquisa estudada Gestão, Política e Manejo Ambiental, com vistas a fornecer ferramentas substanciais à sociedade civil, poder público e gestores tomadores de decisões. Espera-se que os resultados desta pesquisa possam contribuir para direcionar trabalhos futuros que pretendem utilizar técnicas de PLN e servir como modelo para outras instituições avaliarem suas contribuições para a Agenda 2030, utilizando inteligência artificial.

Palavras-chave: Agenda 2030. Sustentabilidade Inteligência Artificial. BERT.

SANTOS, Êmeris Silva. **ODSBAHIA-PTBR**: A natural language processing model to support the 20 sustainable development goals. Advisor: Leonardo Evangelista Moraes. 2025. Dissertation (Master's in Environmental Science & Technology) – Universidade Federal do Sul da Bahia; Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Porto Seguro, 2025.

ABSTRACT

With less than a decade to go until the deadline for achieving the Sustainable Development Goals (SDGs) set by the United Nations (UN) in 2015, there are still gaps in the availability of data and knowledge to monitor progress towards the SDGs and assess the impact of the initiatives implemented. This study investigates how universities in Bahia contribute to the 2030 Agenda, using Natural Language Processing (NLP) techniques. The first chapter presents a theoretical background that explores the concept of sustainable development and the history of the main international conferences that culminated in the formulation of the 2030 Agenda. In addition, the chapter addresses the global and local challenges associated with implementing the SDGs, highlighting the need to integrate the territorial perspective and traditional knowledge. In this context, the fundamental concepts of Artificial Intelligence and their application in SDG-related initiatives are introduced. This approach lays a solid foundation for the development of the technical and analytical contributions presented in subsequent chapters, connecting the 2030 Agenda to the technological tools used in this work The second chapter presents an article that develops ODSBahia-PTBR, an NLP model with deep learning, based on BERTimbau, to classify texts according to the 17 UN SDGs, as well as SDG 18 (Racial Equality), SDG 19 (Art, Culture and Communication) and SDG 20 (Rights of Original Peoples and Traditional Communities), proposed in the 'Agenda 2030 Guide: Integrating SDGs, Education and Society', drawn up in 2020 in partnership between UnB and UNESP. ODSBahia-PTBR achieved a precision of 82%, recall of 72% and F1-score of 77%, in addition to matching 95% of the labels in the OSDG dataset translated into Portuguese, demonstrating its effectiveness in the classification task. Chapter 3 presents descriptive and exploratory research that analyzes data from research and extension projects at the Federal University of Southern Bahia (UFSB), the Federal University of Recôncavo da Bahia (UFRB) and the Federal University of Western Bahia (UFOB). The analysis involves applying the ODSBahia-PTBR to classify the projects according to the SDGs. This research is justified by the importance of understanding, evaluating and directing the research efforts made by higher education institutions and their contributions to achieving the SDGs in their areas of operation. This work is aligned with the PPGCTA, especially the line of research studied Environmental Management, Policy and Stewardship, with a view to providing substantial tools to civil society, public authorities and decision-making managers. It is hoped that the results of this research can contribute to directing future work that intends to use NLP techniques and serve as a model for other institutions to measure their contributions to the 2030 Agenda, using artificial intelligence.

Keywords: Agenda 2030. Sustainability. Artificial Intelligence. BERT.

LISTA DE FIGURAS

Figura 1: Esquema demonstrando as equivalências e transição dos ODM para os ODS. Autor (2024) 20
Figura 2: Abordagens do Aprendizado de Máquina e os principais algoritmos utilizados. Autor (2024). 34
Figura 3: Arquitetura de RNA múltiplas camadas. Fonte: Faceli, 2021..... 39
Figura 4: Modelo da Arquitetura Transformer. Fonte: Autor (2024), adaptado de Vaswani et al., 2017. 42
Figura 5: Estrutura analítica da metodologia do modelo ODSBahia-PTBR. Autor (2024) 52
Figura 6: Distribuição da quantidade de textos por ODS. Autor (2024) 53
Figura 7: Metodologia Analítica para análise das ações universitárias. Autor (2024) 55
Figura 8: Mapa de localização dos três campi da UFSB no território sul da Bahia. O mapa destaca a presença dos campi da UFSB nos municípios de Porto Seguro, Teixeira de Freitas e Itabuna, onde também está a Reitoria. Autor (2024) 57
Figura 9: Mapa de localização da UFRB no Recôncavo baiano. Destaca-se a presença da UFRB em sete municípios: Amargosa, Cachoeira, Feira de Santana, Santo Amaro, Santo Antônio de Jesus, São Félix e tem sua Reitoria localizada em Cruz das Almas. Autor (2024) 59
Figura 10: Mapa de localização da UFOB no Oeste da Bahia, destacando sua presença em cinco municípios. Autor (2024) 60
Figura 11: F1-Score por ODS para cada modelo pré-treinado. Autor (2024) 61
Figura 12: Métricas binárias para cada ODS. Autor (2024) 63
Figura 13: Acurácia por ODS sobre a base OSDG-CD. Autor (2024). 66
Figura 14: Frequência de ocorrências dos ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024) 68
Figura 15: Comparação da frequência de ocorrências dos ODS entre a pesquisa e extensão. Autor (2024) 69
Figura 16: Evolução temporal dos 5 ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024) 70
Figura 17: Grafo de relações entre os ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024) 72
Figura 18: Frequência de coocorrências entre os ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024) .. 73
Figura 19: Participação proporcional de cada universidade entre os ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024) 74
Figura 20:Evolução e Distribuição dos ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023 sob o Tripé da Sustentabilidade. Autor (2024) 75
Figura 21: Proporção das ações universitárias alinhadas aos 5Ps do desenvolvimento sustentável: Pessoas, Planeta, Prosperidade, Paz e Parcerias. Autor (2024) 75
Figura 22: Distribuição dos ODS nos municípios sede das UFSB, UFRB e UFOB. Autor (2024) 76
Figura 23: Distribuição dos ODS nos Projetos de Pesquisa e Extensão (2019-2023) nas UFSB, UFRB e UFOB. Autor (2024) 77

SUMÁRIO

1 INTRODUÇÃO 10
2 OBJETIVOS 15
2.1 Objetivo geral 15
2.2 Objetivos específicos 15
3 REFERENCIAL TEÓRICO 16
3.1 Breve histórico das agendas globais de desenvolvimento 16
3.2 Agenda 2030 e o território: críticas, limites e potencialidades 22
3.3 Como as universidade fazem o que falam sobre a agenda 2030? 26
3.4 Conceitos fundamentais relacionados à Inteligência Artificial 29
3.4.1 Aprendizado de Máquina (AM) 31
3.4.2 Processamento de Linguagem Natural (PLN) 34
3.4.3 Técnicas de Representação Textual 36
3.4.4 Redes Neurais Artificiais (RNAs) 37
3.4.5 Modelos baseados em Atenção 40
3.4.6 Métricas de Avaliação 46
3.5 Integração da Agenda 2030 e a Inteligência Artificial 47
4 METODOLOGIA 51
4.1 Construção do modelo ODSBAHIA-PTBR 51
4.2 Inferência sobre as ações universitárias 55
5 RESULTADOS E DISCUSSÕES 61
5.1 Avaliação do modelo ODSBAHIA-PTBR 61
5.2 Como as universidades fazem o que falam sobre a Agenda 2030? 67
5.3 Limitações da pesquisa 78
6 CONSIDERAÇÕES FINAIS 80
REFERÊNCIAS 83

1 INTRODUÇÃO

Até meados do século XX, o crescimento humano e a expansão das economias mundiais foram amparadas naquilo que o economista Herman E. Daly chamou de modelo do "mundo vazio". Neste modelo, o capital e a força de trabalho gerados pelo ser humano eram tratados como fatores escassos e que, portanto, limitavam o crescimento econômico, enquanto os recursos naturais eram considerados abundantes (Daly, 1991; Daly, 2015). Assim, em linhas gerais, as bases econômicas deste modelo que sustentou a expansão do capitalismo, bem como a reconstrução dos países arrasados pelas duas grandes guerras mundiais, não tiveram uma preocupação ambiental/ecológica (Daly, 1999).

Até esse momento da história, em um intervalo de 110 anos, o mundo testemunhou a população humana saindo da marca de 1 bilhão de pessoas no fim da Revolução Industrial (1850) para cerca de 3 bilhões na década seguinte à Segunda Guerra Mundial. Na sequência, o que se registrou em um intervalo de uma só vida, foi uma nova expansão demográfica para o patamar atual de 7,8 bilhões de pessoas e com a expectativa da marca dos 10 bilhões de pessoas ser alcançada próximo de 2050 (ONU, 2019)[1].

Adicionalmente à explosão demográfica, nos últimos 60 anos, o mundo também testemunhou o crescimento de centros urbanos e o desenvolvimento/expansão das tecnologias da informação e comunicação. Entretanto, estes fenômenos não ocorreram de modo uniforme e acessíveis para todas as pessoas, de modo que, embora a extrema pobreza no mundo venha se reduzindo nas últimas décadas, as desigualdades sociais são cada vez mais intensas (BANCO MUNDIAL, 2019).

Soma-se ainda o fato de que o mais recente relatório do Painel Intergovernamental sobre Mudanças Climáticas informou ao mundo que, entre 2011 e 2020, o aquecimento global atingiu uma média de 1,1 ºC acima dos níveis pré-industriais (1850-1900) (IPCC, 2021). Mais recentemente, a ONU (Organização das Nações Unidas) confirmou que a temperatura média global registrada no segundo semestre de 2023 e ao longo de todo o ano de 2024 foi superior à meta de 1,5 ºC definida pelo Acordo de Paris, celebrado em 2015 por 195 países signatários. Neste ritmo, caso ações rápidas e de grande escala não forem implementadas, o patamar de

[1] Entretanto, desde o final do século XX, muitos países, especialmente os desenvolvidos, passaram a vivenciar uma desaceleração no crescimento, com baixas taxas de natalidade, populações envelhecidas e desafios econômicos relacionados à previdência e à força de trabalho. Em resposta, políticas de incentivo à natalidade têm sido implementadas nesses países para mitigar o envelhecimento populacional e garantir a sustentabilidade dos sistemas previdenciários. (AMAGLOBELI *et al*., 2019; BHATTACHARJEE *et al*., 2021)

aquecimento de 2,0 ºC nas décadas seguintes poderá também estar comprometido (IPCC, 2021; Cheng *et al.* 2025).

É neste cenário paradoxal, repleto de assimetrias, que o mundo também se vê desafiado a fortalecer a compreensão por parte das sociedades de que é preciso ressignificar o conceito de escassez e fazer transições. Em síntese, não obstante o caráter polissêmico do tema, o fato é que a Sustentabilidade deixou de ser um conceito e tornou-se um valor a ser perseguido (Raworth, 2017). A exemplo do que propõe Herman E. Daly, o modelo do "mundo vazio" (ainda vigente) precisa ser substituído pelo modelo do "mundo cheio", a partir do qual os recursos naturais não são mais abundantes e a economia precisa ser compreendida como um subsistema da biosfera finita ao invés de ser tratada como um "ente" independente dos limites do planeta (Daly, 1991; Daly, 2015).

Cabe destacar que há uma vertente crítica ao conceito de desenvolvimento, especialmente a partir do que foi estabelecido no pós segunda guerra mundial com o consenso de Bretton Woods, que definiu o mundo em blocos de países "desenvolvidos" e "subdesenvolvidos" (Santos, 2021). Entretanto, é neste palco histórico das últimas seis décadas que o caminho para a crítica e reflexão sobre os modelos de desenvolvimento vigentes foi pavimentado. A partir daí, ganhou força a compreensão de que desenvolvimento econômico não pode ser visto apenas como sinônimo de crescimento, mas também como resultado qualitativo das condições gerais de vida da população e do meio ambiente.

Esse resultado pode ser exemplificado por iniciativas atuais que promovem não apenas o crescimento econômico, mas também a equidade social e a sustentabilidade ambiental. Por exemplo, programas de transferência de renda, como o Bolsa Família no Brasil, têm demonstrado impactos positivos na saúde e educação das populações mais vulneráveis, reduzindo desigualdades sociais (Jesus, 2025; Martins, 2019). Da mesma forma, iniciativas de manejo sustentável na Amazônia, realizadas por comunidades locais, mostram que é possível combinar geração de renda com preservação da biodiversidade (Miranda *et al.*, 2020). Em países como a Noruega, o equilíbrio entre crescimento econômico e políticas sociais inclusivas reflete-se em altos índices de desenvolvimento humano (IDH) (Grupta, 2024), enquanto na Alemanha, a transição energética para fontes limpas tem contribuído para reduzir emissões de carbono e melhorar a qualidade de vida urbana (Deutschland.de, 2024).

Como desdobramento deste processo, nasceu o conceito de Desenvolvimento Sustentável, bem como todos os produtos frutos dos debates em torno deste conceito (e.g., Agenda 21, Objetivos de Desenvolvimento do Milênio - ODM, Agenda 2030, Objetivos do

Desenvolvimento Sustentável – ODS), e a abordagem contemporânea sobre Sustentabilidade. Apresentado pela primeira vez em 1987, no documento conhecido como Relatório Brundtland, também intitulado “Nosso Futuro Comum” (Our Common Future), esse conceito assume que um compromisso ético e transgeracional da sociedade deve ser adotado para o desenvolvimento das atividades econômicas. Ele assume a compreensão de que o avanço do crescimento econômico de uma sociedade não pode ultrapassar os limites da Biosfera, e não pode comprometer a habilidade das futuras gerações em atender as suas próprias necessidades. Em outras palavras, o conceito de Desenvolvimento Sustentável não diz respeito apenas sobre os impactos das atividades econômicas sobre o meio ambiente, mas também se refere às consequências deles em relação à qualidade de vida e ao bem-estar da sociedade, tanto presente quanto futura.

> “A humanidade tem a capacidade de realizar o desenvolvimento sustentável de modo a garantir as necessidades do presente sem comprometer a capacidade das gerações futuras de atender às suas próprias necessidades. O conceito de desenvolvimento sustentável implica limites, não limites absolutos, mas limitações impostas pelo estado atual da tecnologia e da organização social sobre os recursos ambientais e à capacidade da biosfera de absorver os efeitos das atividades humanas. No entanto, tanto a tecnologia quanto a organização social podem ser gerenciadas e aprimoradas para abrir o caminho para uma nova era de crescimento econômico. A Comissão acredita que a pobreza generalizada não é mais inevitável. A pobreza não é apenas um mal em si, mas o desenvolvimento sustentável requer atender às necessidades básicas de todas as pessoas e estender-se a elas as oportunidades para satisfazer suas aspirações por uma vida melhor. Um mundo em que a pobreza e as desigualdades são endêmicas estará sempre propenso às catástrofes ecológicas dentre outras. “
>
> (Relatório Brundtland, 1987, p. 15)

A Agenda 2030, com seus Objetivos de Desenvolvimento Sustentável (ODS), estabelece um compromisso assumido pelos países membros da Organização das Nações Unidas (ONU) em 2015, para enfrentar desafios históricos e globais, como a erradicação da pobreza, redução das desigualdades, e as mudanças climáticas (ONU, 2015). Na esteira dessa compreensão de que o atual modelo de expansão humana é insustentável ao longo prazo, há uma cobrança importante da sociedade em relação às Instituições de Ensino Superior (IES).

Embora não sejam as únicas responsáveis por enfrentar esses desafios, as universidades desempenham um papel estratégico como produtoras de conhecimento, espaços de pensamento crítico e agentes de integração com a sociedade por meio do ensino, da pesquisa

e da extensão. Outros setores, como governos, empresas, organizações não governamentais e movimentos sociais, compartilham essa responsabilidade de maneira complementar, atuando em conjunto para propor e implementar soluções.

No caso das IES, sua contribuição se destaca no desenvolvimento de ciência, tecnologia e inovação, e na formação de profissionais capacitados, sendo fundamentais para oferecer respostas concretas às demandas de uma sociedade em crise. Portanto, a incorporação da Sustentabilidade nas dimensões essenciais das IES passou a constituir uma condição inegociável de legitimação da sua existência (Marinho, 2014). Assim, conscientes de que a construção de um futuro sustentável deve estar fundamentada em processos que conciliam interesses diversos, fomentem parcerias e explorem soluções inovadoras, as IES vêm respondendo aos desafios deste século (Abramovay, 2012; Raworth, 2017; Zulaf, 2000).

Neste cenário, a Inteligência Artificial (IA) é apresentada como uma importante aliada. Sendo uma das inovações tecnológicas mais significativas, a IA oferece ferramentas avançadas para análise de dados, otimização de processos e tomada de decisões, capazes de contribuir significativamente para a consecução dos ODS (UNESCO, 2019).

Diante da necessidade de monitorar o progresso em direção aos ODS, esta pesquisa investiga como três universidades baianas contribuem para a Agenda 2030, utilizando técnicas de Processamento de Linguagem Natural (PLN). A escolha de usar PLN se deve à sua capacidade de analisar grandes volumes de texto de maneira rápida e precisa, identificando padrões, tendências e informações significativas que seriam difíceis de extrair manualmente.

Como pesquisador, atuo como Analista de Tecnologia da Informação na Universidade Federal do Sul da Bahia (UFSB), uma das instituições analisadas neste estudo. Minha prática profissional está diretamente ligada ao desenvolvimento de soluções tecnológicas, o que me proporciona contato constante com ferramentas para coleta e análise de dados, além de projetos voltados à sustentabilidade e à Agenda 2030. Essa vivência prática, aliada ao interesse acadêmico em temáticas como inovação, desenvolvimento sustentável e tecnologias emergentes, motivou a realização desta pesquisa. A escolha pelo uso de técnicas de PLN surgiu justamente da convergência entre minha experiência na área de TI e o desejo de contribuir, a partir do conhecimento técnico, para o monitoramento e aprimoramento das ações desenvolvidas pelas IES em direção aos ODS.

A presente dissertação parte do reconhecimento de que as universidades enfrentam desafios cada vez mais complexos diante das agendas globais de desenvolvimento. Nesse contexto, torna-se urgente refletir sobre o papel dessas instituições não apenas como centros de

produção científica, mas como agentes estratégicos de transformação social, ambiental e econômica. Ao mesmo tempo, observa-se o surgimento de ferramentas tecnológicas cada vez mais potentes, como a Inteligência Artificial, que abrem novas possibilidades para monitorar, avaliar e qualificar as contribuições das universidades para os ODS.

Este trabalho, portanto, organiza-se em torno de dois eixos complementares: de um lado, a análise crítica da inserção das IES nas agendas globais de desenvolvimento e, de outro, a exploração do potencial de ferramentas de Inteligência Artificial aplicadas à análise de dados institucionais. A partir disso, é apresentado o desenvolvimento do ODSBAHIA-PTBR, um modelo de classificação automática baseado em técnicas de PLN, capaz de identificar o alinhamento de projetos universitários com os ODS.

Para validar esse modelo, foram analisados dados textuais de projetos de pesquisa e extensão registrados entre 2019 e 2023 no Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) das universidades UFSB, UFRB e UFOB. A escolha dessas instituições considerou sua relevância regional, bem como a disponibilidade pública de dados em um mesmo sistema acadêmico, o que possibilitou a coleta automatizada de informações por meio de *web scraping*. Após o pré-processamento dos dados, os textos foram submetidos ao classificador multirrótulo ODSBAHIA-PTBR, permitindo mapear de forma precisa e estruturada a aderência desses projetos à Agenda 2030. Esta análise abrange, inclusive, o ODS 18 (Igualdade Étnico-Racial), incorporado à Agenda nacional, por sua relevância para a realidade brasileira.

Essa abordagem metodológica permitiu sistematizar evidências sobre o papel desempenhado pelas universidades baianas no cumprimento dos ODS, e também demonstrar como tecnologias emergentes podem ser mobilizadas para enfrentar desafios globais de forma mais eficiente e contextualizada.

Este trabalho se alinha ao Programa de Pós-Graduação em Ciências e Tecnologias Ambientais (PPGCTA) da UFSB/IFBA, sobretudo à linha de pesquisa estudada Gestão, Política e Manejo Ambiental, pois está relacionado diretamente a aplicação de tecnologias emergentes para o monitoramento e aprimoramento de práticas sustentáveis, refletindo a necessidade de integrar o conhecimento científico com as ferramentas tecnológicas disponíveis.

Ao longo destas seções, esta dissertação apresenta uma abordagem interdisciplinar ao combinar tecnologia e sustentabilidade que pode ser aplicada e expandida em outros contextos. Espera-se que os resultados desta pesquisa contribuam para gerar e disseminar conhecimento e auxiliar futuras pesquisas ao destacar áreas atendidas e realidades que precisam ser exploradas.

2 OBJETIVOS

2.1 Objetivo geral

Explorar o uso e aplicação de técnicas de Processamento de Linguagem Natural (PLN) na classificação e análise textual de projetos universitários, com foco na identificação das contribuições das universidades baianas para a Agenda 2030.

2.2 Objetivos específicos

- Construir e validar um corpus de treinamento multirrótulo em português, correspondente aos 17 ODS da ONU e mais três novos ODS, para fornecer uma base de dados que ajude a melhorar a avaliação de modelos de classificação de texto.
- Avaliar e comparar o desempenho de modelos pré-treinados em Processamento de Linguagem Natural, a fim de identificar qual deles melhor se adapta ao corpus de treinamento.
- Desenvolver o ODSBahia-PTBR, um modelo de classificação textual ajustado para a língua portuguesa, aprimorando o desempenho na classificação de textos relacionados aos ODS, e avaliá-lo em comparação com outros modelos similares utilizando métricas como precisão (*precision*), *recall* e *F1-score*.
- Mapear as ações de pesquisa e extensão das universidades UFSB, UFRB e UFOB em relação aos ODS utilizando técnicas de Processamento de Linguagem Natural (PLN). Esta análise visa identificar as iniciativas desenvolvidas entre 2019 e 2023, determinar a relação entre essas ações e os ODS, identificar as prioridades e abordagens de cada instituição.

3 REFERENCIAL TEÓRICO

Esta seção aborda a construção da Agenda 2030 e os conceitos fundamentais da Inteligência Artificial (IA), examinando como essas duas áreas, aparentemente distintas, podem se complementar na promoção do desenvolvimento sustentável.

3.1 Breve histórico das agendas globais de desenvolvimento

Desde sua fundação em 1945, a Organização das Nações Unidas (ONU) tem sido um fórum crucial para que seus Estados-membros possam discutir e adotar medidas para enfrentar desafios globais nos âmbitos político, social, ambiental e econômico, através de grandes conferências.

Nas duas primeiras décadas (de 1946 a 1970), as iniciativas da ONU focaram na prevenção de novos conflitos globais, na promoção da paz e na reconstrução de países devastados pela Segunda Guerra Mundial. A Declaração Universal dos Direitos Humanos (1948) foi o seu primeiro documento normativo a proteger os direitos humanos universais e deixou explícito os direitos civis, políticos, econômicos, sociais e culturais básicos que todos os seres humanos devem desfrutar. Ao longo do tempo, ela foi amplamente aceita como as normas fundamentais dos direitos humanos que todos devem respeitar e proteger (ONU, 2008).

Nas décadas de 60 e 70 surgiram inúmeras discussões em torno das preocupações ambientais. Em 1972, o relatório “The Limits to Growth” (1972) produzido por uma equipe multidisciplinar ao Clube de Roma, alertava para os limites físicos do planeta em termos de recursos naturais e capacidade de suporte da biosfera, e argumentava que o crescimento econômico sem limites poderia levar a um colapso irreversível do sistema global (LAGO, 2006).

Nessa época, Meadows *et al*. (1972) analisaram os problemas ambientais e socioeconômicos relacionados ao crescimento desenfreado da população, da produção de alimentos, da industrialização, da poluição e do consumo de recursos naturais não renováveis. Eles argumentaram que todos esses elementos estavam aumentando e que isso estava levando o planeta a uma crise ambiental sem precedentes. Os autores destacaram que o aumento da população estava pressionando os recursos naturais e que a produção de alimentos não estava acompanhando o ritmo do crescimento populacional mundial, o que poderia levar a uma crise de fome. Além disso, a industrialização estava gerando uma enorme quantidade de poluição, afetando a saúde humana e a biodiversidade. (Meadows *et al*., 1972).

Nesse contexto, a questão ambiental ganhou destaque nas conferências da ONU a partir da década de 1970, culminando na Conferência das Nações Unidas sobre o Meio Ambiente Humano realizada em Estocolmo em 1972. Este evento, que ocorreu de 5 a 16 de junho, foi o primeiro grande fórum internacional a destacar a importância da proteção ambiental e seu impacto no desenvolvimento sustentável.

A conferência de Estocolmo reuniu representantes de 113 países que adotaram uma série de princípios para uma gestão sólida do meio ambiente, incluindo a Declaração de Estocolmo e o Plano de Ação para o Meio Ambiente Humano e várias resoluções. A Declaração de Estocolmo trouxe 26 princípios fundamentais e colocou as questões ambientais na vanguarda das preocupações internacionais, marcando o início de um diálogo entre países industrializados e em desenvolvimento sobre a ligação entre o crescimento econômico, a poluição do ar, da água e dos oceanos e o bem-estar das pessoas ao redor do mundo.

Essa primeira conferência levou à criação do Programa das Nações Unidas para o Meio Ambiente (PNUMA) em 1972, concebido para monitorar o estado do meio ambiente, informar a formulação de políticas com ciência e coordenar respostas aos desafios ambientais do mundo (PNUMA, *s.d.*). Por iniciativa da Assembleia Geral da ONU, foi criada a Comissão Mundial sobre o Meio Ambiente e o Desenvolvimento (CMED), em 1983, presidida pela médica e diplomata Gro Harlem Brundtland, que na época era primeira-ministra da Noruega (Ganem, 2012).

Os principais objetivos da CMED eram: 1) reexaminar as questões críticas relativas ao meio ambiente e reformular propostas realísticas para abordá-las; 2) propor novas formas de cooperação internacional nesse campo de modo a orientar as políticas e ações no sentido de fazer as mudanças necessárias, e dar a indivíduos, organizações voluntárias, empresas, institutos e governos uma maior compreensão dos problemas existentes, auxiliando e incentivando-os a uma atuação mais firme (CMMAD, 1991).

Em 1987, essa Comissão apresentou o Relatório Brundtland, documento intitulado “Nosso futuro comum”, que introduziu o conceito de Desenvolvimento Sustentável, definido como "o desenvolvimento que atende às necessidades do presente sem comprometer a possibilidade de as gerações futuras atenderem a suas próprias necessidades" (CMMAD, 1991). Este conceito surgiu como uma terceira via para mediar dois extremos no debate sobre desenvolvimento econômico e ambiental: de um lado, desenvolvimentistas que defendiam o desenvolvimento econômico contínuo e, de outro, defensores do crescimento zero que viam o crescimento ilimitado como insustentável (Romeiro, 2012).

Embora o Relatório Brundtland seja frequentemente citado por trazer a definição do desenvolvimento sustentável, o documento trouxe também outros tópicos afins, como o papel da economia internacional, população e recursos humanos, segurança alimentar, espécies e ecossistemas, energia, indústria e princípios legais propostos para proteção ambiental. Foram apontados no relatório dois impedimentos para o progresso global: a pobreza nos países de Terceiro Mundo e o consumismo excessivo nos países de Primeiro Mundo. O relatório criticou a extrema pobreza presente, principalmente, nos países em desenvolvimento, que limita o acesso a recursos básicos como saúde, educação e alimentação, forçando as populações mais pobres a explorar os recursos naturais de maneira insustentável, degradando o meio ambiente em busca de subsistência. Além disso, o relatório também critica os padrões de consumo insustentáveis nos países desenvolvidos e o uso desproporcional dos recursos naturais contribuindo para a degradação ambiental global (Jarvie, 2016).

Em 1992 no Rio de Janeiro, a Conferência sobre Meio Ambiente e Desenvolvimento, ou Cúpula da Terra, também conhecida como Rio 92 (ou ECO 92), aprofundou e expandiu as discussões iniciadas em Estocolmo. Realizada de 3 a 14 de junho, a conferência contou com a participação de 175 países e produziu importantes documentos para a agenda ambiental global. Entre eles estavam tratados globais que buscam enfrentar as mudanças climáticas e a perda de biodiversidade; a Declaração do Rio sobre Meio Ambiente e Desenvolvimento, um conjunto de 27 princípios que delinearam os direitos e responsabilidades das nações em relação à sustentabilidade ambiental; e a Agenda 21, um plano de ação abrangente para promover o desenvolvimento sustentável a nível internacional, nacional e local. Nessa conferência o conceito de desenvolvimento sustentável foi consolidado como uma meta central das políticas internacionais (Lago, 2006).

Nos anos 2000, a Cúpula do Milênio, também realizada pela ONU, reuniu líderes de 191 países em Nova Iorque, de 6 a 8 de setembro de 2000, com o objetivo de debater os principais problemas globais e estabelecer metas para o desenvolvimento. Nesta ocasião, foi assinada a Declaração do Milênio das Nações Unidas (Resolução nº 55/2 da Assembleia Geral da ONU), que incluiu os seguintes Objetivos de Desenvolvimento do Milênio (ODM): (1) Acabar com a fome e a miséria; (2) Oferecer educação básica de qualidade para todos; (3) Promover a igualdade entre os sexos e a autonomia das mulheres; (4) Reduzir a mortalidade infantil; (5) Melhorar a saúde das gestantes; (6) Combater a Aids, a malária e outras doenças; (7) Garantir qualidade de vida e respeito ao meio ambiente; (8) Estabelecer parcerias para o desenvolvimento (ONU, 2000).

Os ODM representaram uma mudança na abordagem global para o desenvolvimento, focando em metas quantificáveis e específicas a serem alcançadas até 2015, funcionando como catalisadores para a ação, aumento da cooperação internacional e incentivando os países a incorporarem as 21 metas em suas políticas nacionais de desenvolvimento. Um conjunto de 60 indicadores foram utilizados para o acompanhamento de progresso dos ODMs.

Muitos países incorporaram os ODM em suas estratégias de desenvolvimento e implementaram medidas para alcançar essas metas. Contudo, o progresso foi desigual e, apesar dos esforços, muitos países não cumpriram uma ou mais metas (ONU, 2013). No Brasil, a governança dos ODM ficou a cargo do "Grupo Técnico para Acompanhamento das Metas e Objetivos de Desenvolvimento do Milênio", instituído pelo Decreto Presidencial de 31 de outubro de 2003. Esse Grupo Técnico (GT) foi responsável por “nacionalizar” os ODM, as metas e o conjunto de indicadores a elas associados à realidade brasileira, conforme prática incentivada pela ONU a todos os países (IPEA, 2004).

O Brasil apresentou bons resultados com os ODM, alcançando muitas das metas estabelecidas e em algumas chegou a superar as metas globais, como a erradicação da extrema pobreza (ODM 1) e a redução da mortalidade infantil (ODM 4). Apesar dos sucessos que o Brasil alcançou em várias áreas dos ODM, algumas metas não foram plenamente alcançadas até 2015, como melhorar a saúde materna (ODM 5) e promover a igualdade de gênero e autonomia das mulheres (ODM 3) (Roma, 2013).

Em 2012, faltando três anos para atingir o prazo estabelecido para alcançar os ODM, o Rio de Janeiro sediou mais uma conferência, a Rio+20, com o objetivo de reavaliar o progresso e renovar o compromisso político global com o desenvolvimento sustentável. Realizada de 20 a 22 de junho, a Rio+20 focou em dois temas principais: a economia verde no contexto do desenvolvimento sustentável e da erradicação da pobreza, e a estrutura institucional para o desenvolvimento sustentável. A conferência resultou no documento "O Futuro que Queremos", que delineia uma visão e metas para promover um desenvolvimento sustentável global, refletindo o avanço e as novas prioridades emergentes desde a Eco-92.

Todas essas conferências realizadas, bem como os seus artefatos, ajudaram a construir uma institucionalidade internacional, abrindo um caminho comum para negociação e cooperação entre as nações. Como um desdobramento dessas iniciativas e para continuar o esforço global em direção a um desenvolvimento sustentável mais abrangente e inclusivo, após mais de dois anos de negociações, os 193 países-membros integrantes da Assembleia Geral da ONU, adotaram, em 25 de setembro de 2015, uma nova agenda para substituir os ODM.

Além desses chefes de Estado, vários representantes da sociedade civil e com a participação de milhares de pessoas de todo o planeta por meio da Plataforma My World[2], idealizaram e acordaram o documento "Transformando nosso mundo: a Agenda 2030 para o Desenvolvimento Sustentável", visando o projeto maior que é o Desenvolvimento Sustentável Global e a partir de uma transição dos ODMs para o que passou a ser chamado de Objetivos do Desenvolvimento Sustentável (ODS) (Figura 1) (ONU, 2015).

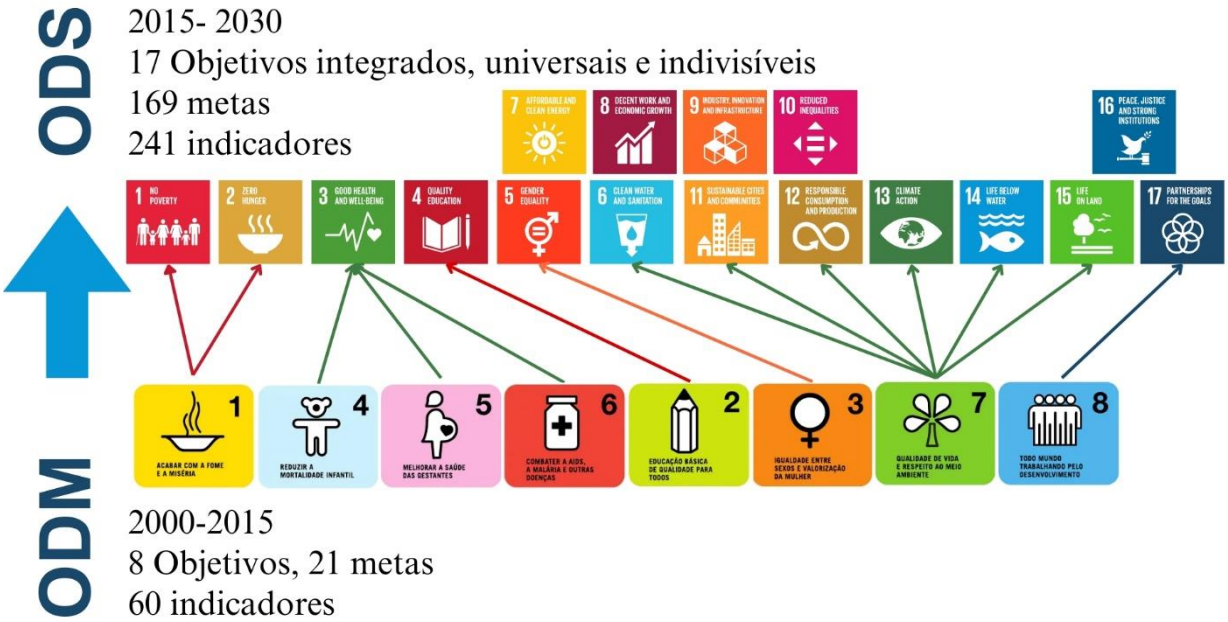

***Figura 1**: Esquema demonstrando as equivalências e transição dos ODM para os ODS. Autor (2024)*

A Agenda 2030 da ONU estabelece 17 ODS, acompanhados de 169 metas e 241 indicadores para monitoramento. Ela orienta políticas nacionais e cooperação internacional, sucedendo os ODM. A iniciativa busca equilibrar as dimensões econômica, social e ambiental, promovendo a erradicação da pobreza, a proteção do planeta e o bem-estar global. Seus objetivos são agrupados em cinco áreas-chave: pessoas (ODS 1 a 6), prosperidade (ODS 7 a 12), planeta (ODS 13 a 15), paz (ODS 16) e parcerias (ODS 17) (Morton *et al*., 2017).

Como signatário da ONU, o Brasil aderiu à Agenda 2030 e se comprometeu a implementá-la por meio de políticas públicas e estratégias nacionais. Para isso, o país realizou a internalização da Agenda, adaptando suas metas e indicadores à realidade brasileira, incluindo aspectos econômicos, sociais e ambientais.

[2] Disponível em: https://myworld2030.org/?lang=pt. Acesso em: 28 ago. 2024.

Em 2016, foi estabelecida uma governança nacional ao processo de implantação da Agenda 2030 no país, criando a Comissão Nacional para os ODS (CNODS), um colegiado de natureza consultiva no âmbito da Secretaria-Geral da Presidência da República cujo objetivo é contribuir para a internalização da Agenda 2030 no país, estimular a sua implementação em todas as esferas de governo e junto à sociedade civil, além de acompanhar, difundir e dar transparência às ações realizadas para o alcance das suas metas e ao progresso no alcance dos ODS (CNODS, 2023).

O Instituto de Pesquisa Econômica Aplicada (IPEA) foi o responsável pela adequação das metas globais dos ODS para o Brasil. O Instituto também produziu um retrato da situação brasileira para alcançar as 169 metas da Agenda 2030, compilado na série Cadernos ODS[3] (IPEA, 2024). Outro órgão que também desempenha um papel fundamental nesse contexto é o Instituto Brasileiro de Geografia e Estatística (IBGE), por meio da plataforma ODS Brasil[4], que busca apresentar indicadores atualizados para monitorar o progresso das metas no país (IBGE, 2024).

Além das análises realizadas por órgãos governamentais, organizações da sociedade civil têm se mobilizado para acompanhar e avaliar o progresso das principais metas globais para o Brasil, a exemplo do Grupo de Trabalho da Sociedade Civil para a Agenda 2030 (GT Agenda 2030).

O GT Agenda 2030 foi estabelecido em 9 de setembro de 2014, fruto da colaboração entre organizações não governamentais (ONGs), movimentos sociais e fundações brasileiras, para promover e monitorar a implementação da Agenda 2030 no Brasil. Com cerca de 50 membros de diferentes setores, o grupo foca em difundir os ODS, mobilizar a sociedade civil, e influenciar políticas junto ao governo brasileiro e à ONU.

Uma das importantes contribuições deste GT é a produção anual do Relatório Luz[5], que desde 2017 tem analisado a implementação dos ODS no país e mostrado o que é preciso fazer para cumprir o compromisso que assumiu junto à ONU de alcançar as metas globais até 2030. A contribuição desta atuação do GT Agenda 2030 foi especialmente importante no período de 2016 a 2022, quando os governos da época desmobilizaram as estruturas de suporte de implementação e acompanhamento da Agenda 2030 no Brasil.

---

[3] Disponível em: https://www.ipea.gov.br/ods/publicacoes.html Acesso em 28 ago 2024.

[4] Disponível em: https://odsbrasil.gov.br/ Acesso em 28 de set 2024

[5] https://gtagenda2030.org.br/biblioteca/relatorios-luz/

Em 20 de dezembro de 2023, o Brasil deu um passo histórico ao criar o ODS 18 (Igualdade Étnico-Racial), focado na promoção da igualdade étnico-racial. Esse ODS visa eliminar o racismo e a discriminação étnico-racial contra povos indígenas, afrodescendentes e grupos populacionais afetados por múltiplas formas de discriminação. Por meio da Resolução nº 02/CNODS, foi instituída a Câmara Temática sobre Igualdade Étnico-Racial, reforçando o compromisso do país em enfrentar desigualdades históricas e garantir inclusão social para todas as raças e etnias. A construção das metas e indicadores do ODS 18 começou em 2024, com um processo participativo envolvendo governo, movimentos sociais, academia e setor privado. Esse esforço está sendo liderado pelo Ministério da Igualdade Racial e o Ministério dos Povos Indígenas, em parceria com o BNDES, para implementar políticas públicas que promovam justiça social e oportunidades iguais para todos (Brasil, 2023).

3.2 Agenda 2030 e o território: críticas, limites e potencialidades

A proposta desta seção é explorar algumas reflexões acerca da Agenda 2030 sob uma perspectiva territorial, analisando como os ODS se comportam quando confrontados com as realidades específicas dos territórios locais. Conforme apresentado anteriormente, ao longo dos últimos anos, a Agenda 2030 tem se consolidado como um marco global para o desenvolvimento sustentável, oferecendo um conjunto de metas e indicadores aplicáveis a todos os países. No entanto, sua implementação em contextos locais muitas vezes revela alguns desafios, incluindo a falta de sensibilidade às particularidades culturais, sociais e ambientais de cada região.

Apesar de seu caráter abrangente e intenções positivas, a implementação dessa Agenda tem enfrentado críticas substanciais, especialmente quando observada sob a ótica do conceito de desenvolvimento sustentável introduzido pelo Relatório de Brundtland (1987), na tentativa de conciliar crescimento econômico e proteção ambiental.

Para alguns autores, esse conceito emergiu em resposta à crise ambiental global e foi consolidado por organismos como a ONU, mas tornou-se um pilar do discurso empresarial e governamental, sem, contudo, questionar as premissas fundamentais do capitalismo. Adicionalmente, Santos (2019) alerta sobre o papel do discurso político em torno do conceito de desenvolvimento no pós-segunda guerra, que dividiu o mundo em dois blocos, países desenvolvidos e subdesenvolvidos, resultando como ferramenta de imposição de uma nova ordem mundial que subalternizou os países mais pobres economicamente.

Para Vizeu *et al*. (2012), o desenvolvimento sustentável é essencialmente ideológico, pois cria a ilusão de que é possível conciliar preservação ambiental e crescimento econômico. Eles sustentam que essa noção mascara a destruição ambiental e social, legitimando a continuidade do sistema capitalista. Para os autores, uma mudança real exigiria um rompimento com essa lógica, priorizando o ser humano e a coletividade em vez do lucro (Vizeu *et al*., 2012).

Nessa mesma perspectiva, Kruse & Cunha (2022) reforçam essa crítica, propondo que se rompa com a dependência do mercado e valorize a cooperação entre os países do Sul Global, além da adoção de políticas que priorizem a sustentabilidade real, e não apenas medidas paliativas que mascaram os danos ecológicos e sociais causados pelo capitalismo. Em linhas gerais, para além de atuar nas consequências das crises sociais e ambientais, é necessário também atuar em suas causas.

No contexto da Agenda 2030, essa crítica se torna ainda mais relevante, pois a implementação dos ODS frequentemente ignora as contradições do modelo econômico vigente, tratando a transição ecológica como um processo conciliável com a lógica da acumulação de capital. Essas reflexões reforçam a necessidade de uma análise crítica sobre como os ODS são aplicados nos diferentes territórios, especialmente nos mais vulneráveis, onde os efeitos da degradação ambiental e da desigualdade se manifestam de forma mais expressiva (Favareto, 2020).

Diante desse cenário, torna-se fundamental considerar a territorialização como um elemento-chave para a adaptação e efetivação das políticas de desenvolvimento sustentável. Conforme discutido por Froio & Gonçalves (2024), a territorialização é um processo político essencial para a governança e implementação de políticas públicas, pois envolve não apenas a transformação de espaços físicos, mas também a construção de elementos políticos, econômicos e simbólicos. No Brasil, embora a União Federal exerça um papel estratégico na formulação dessas políticas, são os governos subnacionais, especialmente estados e municípios, que assumem um papel central na implementação e adaptação das ações às realidades locais.

Nesse sentido, a territorialização torna-se um mecanismo fundamental para a implementação da Agenda 2030 e os ODS, pois está diretamente ligada ao processo de localização dos ODS — ou seja, à adaptação das metas globais às especificidades e necessidades locais. Ao serem formulados como metas globais, os ODS nem sempre dialogam diretamente com as realidades locais, exigindo um processo de adaptação territorial para sua efetivação. Dessa forma, governos subnacionais (estaduais e municipais) assumem um papel essencial na definição de estratégias, metas e ações concretas para alcançar os objetivos

estabelecidos pela ONU (Froio; Gonçalves, 2024). A centralidade dos governos locais na implementação da Agenda 2030 se justifica ao fato de estarem mais próximos das populações e comunidades, o que lhes permite compreender melhor as demandas específicas de cada território e atuar de forma mais eficaz.

Conforme analisado por Ramiro (2024), a implementação da Agenda 2030 no Brasil foi marcada por ciclos políticos que influenciaram diretamente seu avanço. Diante da falta de coordenação e apoio do governo federal, especialmente entre 2018 e 2022, governos subnacionais e organizações da sociedade civil assumiram um papel central na adaptação dos ODS às realidades locais.

Nesse período, o Brasil viveu um paradoxo na territorialização dos ODS: enquanto o governo federal retrocedeu devido ao desmonte institucional e à retórica populista de direita alinhada a uma agenda neoliberal, estados e municípios avançaram. Graças à descentralização administrativa, esses governos subnacionais resistiram, legislando e implementando políticas sustentáveis, muitas vezes em oposição ao governo central (Froio; Gonçalves, 2024).

Para mitigar a ausência de uma estratégia nacional, estados e municípios se articularam por meio de redes locais como a Confederação Nacional de Municípios (CNM) e transnacionais como a Governos Locais pela Sustentabilidade (ICLEI), incorporando os ODS em seus planos de desenvolvimento (Ramiro, 2024). Esse cenário reforça a importância da autonomia local na manutenção dos ODS, mesmo em contextos políticos adversos. Além disso, a territorialização da Agenda 2030 no Brasil, não ocorreu apenas por meio dos governos locais, mas também pela resistência da sociedade civil, que manteve o debate sobre os ODS vivo e pressionou por políticas públicas alinhadas ao desenvolvimento sustentável.

Como destacado por Nilo (2024), a sociedade civil brasileira exerceu um papel fundamental no monitoramento e na implementação. Durante o governo Bolsonaro (2019-2022), a Comissão Nacional dos ODS (CNODS) foi extinta, e a Agenda 2030 foi excluída do Plano Plurianual (PPA 2020-2023), evidenciando a falta de compromisso do governo federal com os objetivos globais. Nesse contexto, organizações como a Gestos e a Abong mobilizaram redes nacionais e internacionais para garantir que os ODS continuassem a ser discutidos e aplicados em diferentes territórios, mesmo sem apoio estatal (Nilo, 2024).

A atuação de ONGs e movimentos sociais, conforme analisado por Nilo (2024), não se limitou ao contexto nacional, mas também influenciou negociações internacionais e debates multilaterais da ONU. Esse envolvimento possibilitou a inclusão de pautas progressistas, como

a redução das desigualdades (ODS 10), e garantiu que a participação da sociedade civil fosse reconhecida no monitoramento e revisão da Agenda 2030.

No Brasil, essa resistência também se materializou em iniciativas como a produção do Relatório Luz, apresentado na seção anterior, um relatório anual alternativo que avalia o progresso dos ODS no país e foi fundamental diante dos retrocessos institucionais. Essa atuação da sociedade civil demonstra que a territorialização da Agenda 2030 não pode ser vista apenas como um processo conduzido pelo Estado, mas sim como um esforço coletivo, onde redes locais, academia, movimentos sociais e organismos internacionais desempenham um papel complementar na adaptação das metas globais às realidades locais (Nilo, 2024).

É nessa perspectiva de território que Gehre *et al*. (2024) enfatizam a necessidade de formação de redes colaborativas entre universidades, potencializando suas capacidades na localização da Agenda 2030 por meio da articulação entre ensino, pesquisa e extensão. As universidades são retratadas como microcosmos dos territórios locais, contribuindo diretamente para economias regionais e promovendo soluções para problemas cotidianos, como mobilidade urbana, resíduos sólidos e práticas sustentáveis.

Nesse contexto de colaboração entre universidades para a territorialização da Agenda 2030, Cabral e Gehre (2020) propõem a inclusão de três novos ODS, ampliando a abordagem dos desafios globais para contemplar realidades e grupos historicamente invisibilizados. Essa proposta foi formalizada no documento 'Guia Agenda 2030: Integrando ODS, Educação e Sociedade', elaborado em 2020 em parceria entre a UnB e a UNESP, destacando temas que, embora não explicitamente representados nos 17 ODS, possuem grande relevância para o Brasil e o mundo. Essa proposta que inclui "Igualdade Racial (ODS 18)", "Arte, Cultura e Comunicação (ODS 19)", e "Direitos dos Povos Originários e Comunidades Tradicionais (ODS 20)", visa dar visibilidade a grupos historicamente invisibilizados e negligenciados nas agendas globais de desenvolvimento, garantindo que aspectos importantes para o desenvolvimento e sustentabilidade locais sejam abordados (Cabral; Gehre, 2020).

A inclusão desses novos objetivos representa um exercício crítico de reimaginar e ressignificar as caracterizações da Agenda 2030, reforçando seu compromisso com a inclusão no desenvolvimento sustentável. Embora não integrem oficialmente a Agenda 2030 da ONU ou sua versão adotada no Brasil, esses três ODS complementam os existentes e ampliam a abordagem sobre as demandas sociais atuais.

Embora a luta contra o racismo esteja transversalmente incluída em outros ODS, como a erradicação da pobreza, a garantia de educação de qualidade, e a redução das desigualdades,

o ODS 18 traz à evidência essa problemática. Ele busca promover a igualdade racial a partir do enfrentamento de todos os tipos de racismo, destacando sua importância para a construção de um desenvolvimento sustentável e inclusivo. Já o ODS 19 destaca o impacto social da arte, promovendo a mobilização de grupos sociais, novas percepções de mundo, e a construção de conhecimento sobre as dinâmicas sociais. O ODS 20, por sua vez, visa preservar e valorizar a cultura e o conhecimento tradicional, garantindo os direitos de povos indígenas, quilombolas, ribeirinhos e outros grupos sociais.

3.3 Como as universidade fazem o que falam sobre a agenda 2030?

Esse olhar crítico sobre a abrangência e os desafios da Agenda 2030 discutidos anteriormente, reforça o papel estratégico das universidades na promoção do desenvolvimento sustentável em seus territórios de atuação. As universidades oferecem aos indivíduos a oportunidade de desenvolver seu potencial humano de forma integral, independentemente de classe social, etnia e gênero (Chankseliani; McCowam, 2021), e são essenciais na geração de conhecimento científico e tecnológico, bem como na formação de pessoas capacitadas e conscientes das questões ambientais, sociais e econômicas (Silva; Almeida, 2019; Savegnago *et al*., 2022). Ao promoverem ações e estratégias baseadas em evidências e conduzir pesquisas, elas viabilizam a consecução dos ODS (Bendin; Faria, 2021; Allen *et al*., 2021). Elas.

Para Leal Filho *et al*. (2019) quando as universidades conseguem demonstrar o impacto dos ODS na resolução de problemas do mundo real relacionados ao desenvolvimento sustentável, elas se tornam mais relevantes aos olhos do governo, da comunidade local e dos parceiros de cooperação. Além disso, elas estão sendo cada vez mais cobradas a demonstrar seu impacto na sociedade, e nesse contexto, as publicações acadêmicas de pesquisas que possuem relação com os objetivos e metas da Agenda 2030 são consideradas uma unidade adequada para fundamentar essas avaliações (Purnell, 2022).

Apesar do papel estratégico das universidades na implementação dos ODS, a mensuração dessas contribuições ainda enfrenta desafios metodológicos, devido a lacunas de dados e conhecimento no rastreamento de políticas e contribuições, que são essenciais para monitorar o progresso em direção aos ODS e para a avaliação do impacto das iniciativas implementadas (Sachs *et al*., 2019).

Por exemplo, em sua revisão Alcántara-Rubio *et al*. (2022) identificaram que uma das ações menos frequentes entre as universidades está em “capacitar estudantes na pesquisa dos ODS” e “fomentar e promover os ODS como linha de pesquisa dentro da universidade”. Essa

falta de ênfase na capacitação acadêmica pode refletir uma das razões pelas quais há dificuldades em coletar dados que suportem uma avaliação mais robusta das iniciativas das universidades em relação aos ODS.

Contudo, as universidades desempenham um papel multifacetado na promoção e implementação da Agenda 2030, estando envolvida em uma ampla gama de atividades que perpassa o ensino, pesquisa e seu envolvimento com a sociedade, por meio de transferência de conhecimento, desenvolvimento comunitário e engajamento público (Fia *et al*., 2022). Além disso, estão adotando práticas para gerir suas operações e estruturas internas de maneira mais sustentável, reduzindo seu impacto ambiental e promovendo a equidade social e econômica (Leal Filho *et al*., 2021; Torres *et al*., 2020).

De todo modo, as universidades públicas brasileiras têm impulsionado inovações institucionais para a implementação dos ODS, ajustando sua estrutura e estratégia para internalizar a Agenda 2030. Essas inovações incluem a criação de comissões, projetos e órgãos administrativos específicos para coordenar ações relacionadas aos ODS; a reformulação de normas institucionais para alinhar missão e atividades acadêmicas aos princípios da sustentabilidade; e o estímulo à integração entre ensino, pesquisa e extensão, reforçando a curricularização dos ODS no ambiente universitário. Essas iniciativas fortalecem o seu papel como agentes transformadores no desenvolvimento sustentável (Galvão *et al*., 2024).

Estudos recentes investigaram as contribuições das universidades para a Agenda 2030 e os ODS. Savegnago *et al*. (2022) analisaram os Planos de Desenvolvimento Institucional (PDIs) de 67 universidades federais brasileiras, identificando estratégias adotadas para incorporar os ODS em suas atividades acadêmicas e administrativas. Inácio *et al*. (2024) avaliaram o desempenho das universidades brasileiras no alcance dos ODS entre 2020 e 2024, utilizando o *University Impact Rankings* como referência. Os resultados indicaram um aumento na participação das instituições brasileiras no ranking, com destaque para os ODS 3 (Saúde e Bem-Estar), 4 (Educação de Qualidade) e 8 (Trabalho Decente e Crescimento Econômico).

Costa e Silva *et al.* (2024), analisaram a institucionalização da Agenda 2030 em universidades da Amazônia Legal brasileira, com base na análise de documentos institucionais e revisão bibliométrica da literatura internacional sobre o tema. Os resultados indicaram que essas universidades ainda enfrentam desafios para fomentar a implementação dos ODS, embora desempenhem um papel central na promoção do ensino, pesquisa e extensão voltados à sustentabilidade, especialmente em contextos de vulnerabilidade socioambiental.

O Observatório de Sustentabilidade na Educação Superior da América Latina e Caribe (Oses-ALC) lançou, em 2022, um formulário de diagnóstico para avaliar o conhecimento, compromisso e ações das Instituições de Ensino Superior (IES) brasileiras em relação à Agenda 2030. Os resultados preliminares indicam um avanço significativo no engajamento das IES com os ODS, embora haja variações regionais e institucionais (OSES-ALC, 2024).

Além disso, pesquisas recentes têm explorado o uso de técnicas de Inteligência Artificial (IA) para analisar a contribuição das universidades para os ODS. Entre essas técnicas, destaca-se o Processamento de Linguagem Natural (PLN), uma subárea da IA que combina conhecimentos de linguística, ciência da computação e aprendizado de máquina para processar, interpretar e gerar linguagem natural (Faceli *et al*., 2021). No contexto da Agenda 2030, o PLN possibilita, por exemplo, a extração e análise automatizada de grandes volumes de textos institucionais e acadêmicos identificando padrões semânticos e tendências associadas aos ODS.

Na Universidade Federal de Uberlândia (UFU), um estudo aplicou técnicas de PLN para mapear as ações de extensão da instituição e verificar seu alinhamento com os ODS. A análise de mais de 15 mil registros (programas, projetos, eventos e cursos) desenvolvidos entre 2009 e 2022 revelou que a maioria das ações estava alinhada às dimensões social e econômica, seguidas pelas dimensões ambiental e de parcerias (Borsatto *et al*. 2022).

Outro estudo na UFU analisou a contribuição das pesquisas de pós-graduação para os ODS, utilizando técnicas de IA, como modelagem de tópicos. A análise de 2.563 dissertações e teses desde 2015 indicou a predominância dos ODS 3, 4, 9, 8 e 2, relacionados à saúde, educação, inovação, crescimento econômico, trabalho digno, agricultura sustentável e fome zero, respectivamente (Ribeiro, 2024). Por fim, uma revisão sistemática da literatura aplicou ferramentas de IA para avaliar como os ODS estão sendo abordados no contexto universitário. A análise de 97 artigos constatou um crescimento no interesse acadêmico pela integração dos ODS nas universidades (Ribeiro *et al*., 2022).

Apesar dessas investigações, ainda há uma lacuna quanto à análise regionalizada das contribuições das universidades baianas. Nenhum dos estudos mencionados foca especificamente nesse contexto, reforçando a necessidade de aprofundamento dessa temática, que será abordada neste trabalho por meio da análise de ações institucionais e seu alinhamento com os ODS.

Entender como as universidades fazem o que falam sobre a Agenda 2030 é fundamental, pois revela não apenas a efetividade das ações e compromissos assumidos pelas

IES em relação aos ODS, mas também destaca as lacunas e oportunidades que podem existir na sua implementação.

Este estudo concentra-se na análise de projetos de pesquisa e ações de extensão em universidades baianas, investigando sua contribuição para a Agenda 2030. Esse foco se justifica pela importância dessas atividades na inovação e transformação social, bem como pela necessidade de avaliar como as universidades operacionalizam seus compromissos institucionais em ações concretas que territorializam os ODS.

Ao promover a interação entre academia e comunidades locais, a pesquisa e a extensão, elas fortalecem a relação com a sociedade e contribuem diretamente para o desenvolvimento regional (Galvão *et al*., 2024). Identificar os relacionamentos das ações universitárias com os ODS, constitui-se um passo importante para a criação de políticas e estratégias eficazes para o desenvolvimento sustentável.

Além disso, é importante destacar que a abordagem presente nesta pesquisa pode ser particularmente relevante para instituições de ensino mais jovens, que, muitas vezes, enfrentam restrições de recursos significativas. Isso porque, universidades maiores (em termos de estudantes matriculados), mais antigas e mais diversas (tanto em termos de diversidade internacional como de diversidade de gênero) já possuem uma tendência maior de comunicar seu compromisso com os ODS em suas atividades (De Iório *et al*., 2022).

3.4 Conceitos fundamentais relacionados à Inteligência Artificial

No contexto da sustentabilidade global, a Agenda 2030 é o que temos como o documento mais abrangente e atual, orientando as ações em direção a um desenvolvimento sustentável, inclusivo e equilibrado. De forma análoga, no campo das inovações tecnológicas, a Inteligência Artificial (IA) se destaca como uma das realizações mais significativas, posicionada no ápice do desenvolvimento tecnológico da raça humana, refletindo o progresso científico e potencial desta tecnologia que pode trazer contribuições significativas para a implementação desta Agenda (Medeiros, 2018).

Os termos 'inteligência artificial' carregam em cada elemento compreensões diversas de seus significados, podendo ser analisados juntos ou separadamente, sobretudo, nas discussões filosóficas e científicas sobre o que constitui a complexidade da inteligência humana (Menezes, 2023). Afinal, como a “inteligência”, produto da atividade cerebral e mental do ser humano, pode se manifestar fora dele? No entanto, é importante ressaltar que este capítulo

apenas traz luz aos conceitos fundamentais que norteiam esta pesquisa, não adentrando profundamente nessas reflexões.

Para conceituar Inteligência Artificial (IA), é essencial primeiro definir 'inteligência', termo abordado por várias áreas e pesquisadores. Conforme o psicólogo Howard Gardner (1995), a inteligência consiste na habilidade de resolver problemas ou criar produtos que sejam significativos em um ou mais ambientes culturais. Essa definição enfatiza a diversidade de formas de inteligência, sugerindo que a inteligência não é uma única capacidade, mas sim um conjunto de habilidades distintas que se manifestam em diferentes áreas.

Para o psicólogo cognitivo Robert Sternberg (2010), é a “capacidade que o ser humano tem para aprender com a experiência”. E para Jean Piaget, estudioso dos processos de construção do conhecimento humano, significa “Interações do indivíduo com o ambiente, envolvendo um equilíbrio entre a assimilação e acomodação” (Lefrançois, 2013). Segundo o pesquisador de IA Marvin Lee Minsky (1985), "inteligência" pode ser vista como qualquer processo mental que nos permite resolver problemas difíceis, especialmente aqueles processos que ainda não compreendemos.

Dessa forma, tem-se a inteligência como sendo uma capacidade mental geral que, entre outras coisas, envolve a habilidade de raciocinar, planejar, resolver problemas, pensar de forma abstrata, compreender ideias complexas, aprender rapidamente e adquirir conhecimento com a experiência (Gottfredson, 1997).

Conforme Wang (2019), não há uma maneira única de conceituar a IA, por existir uma série de elementos que se manifestam de maneira diferenciada e também quanto às diferentes interpretações em como os processos de IA se correlacionam com os mecanismos do cérebro e da mente humana. Mas é inevitável uma definição funcional. Na concepção do autor, todas as definições de IA podem ser vistas como “abstrações da inteligência humana, correspondendo a racionalidade a um determinado nível”.

Na definição de Minsky (1980), a IA é “a ciência que permite fazer com que as máquinas realizem tarefas que necessitariam de inteligência, se elas fossem efetuadas pelos homens”. Ou seja, máquinas capazes de realizar atividades complexas, imitar ou reproduzir aspectos da inteligência humana por meio de algoritmos e processos computacionais. Da mesma forma, na concepção de John McCarthy (2007), criador desse termo 'Inteligência Artificial' em 1956, a IA é definida como 'a ciência e a engenharia de fabricação de máquinas inteligentes, especialmente programas de computador inteligentes'

Russell e Norvig (2022), por sua vez, categorizam as definições de Inteligência Artificial usando duas distinções principais: 'humano versus racional' e 'pensar versus agir'. A partir dessas distinções, eles trazem quatro abordagens principais: sistemas que pensam como humanos, sistemas que pensam de forma racional, sistemas que agem como humanos e sistemas que agem de forma racional. Com base na categorização proposta por esses autores, sistemas que pensam estão relacionados aos processos de raciocínio (cognitivos ou racionais), enquanto sistemas que agem estão voltados para o comportamento observável, seja humano ou racional.

Ao longo da história, a abordagem acerca de sistemas que agem racionalmente, segundo Russel e Norvig (2022), tem prevalecido sobre as demais. Os autores definem um agente inteligente como tudo o que pode ser considerado "capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por intermédio de atuadores". O agente inteligente pode ser reativo, cujas ações são plenamente determinadas pelo seu criador, ou baseados em aprendizagem, capazes de ajustar seu comportamento com base em experiências passadas (Russel; Norvig, 2022).

De fato, os sistemas de IA eram baseados em operações lógicas buscando alcançar determinado fim, por meio de regras explícitas. Atualmente, observa-se a utilização de métodos baseados em probabilidades e aprendizado de máquina (*Machine Learning*), uma abordagem da IA que se concentra no desenvolvimento de algoritmos e modelos capazes de aprender padrões a partir de dados.

Essa transformação permitiu que os sistemas desenvolvessem a capacidade de tomar decisões com base em padrões identificados e experiências passadas. Atualmente, a IA é vista como um assunto interdisciplinar que envolve informação, lógica, biologia, psicologia, filosofia e muitas outras disciplinas, e tem sido usada para a construção de robôs inteligentes (*hardwares/softwares*), reconhecimento de fala, processamento de imagem e processamento de linguagem natural, objeto deste estudo. (Zhang; Lu, 2021).

3.4.1 Aprendizado de Máquina (AM)

Nas primeiras décadas da IA, especialmente entre 1970 e 1980, muitos problemas eram tratados com Sistemas Especialistas (ou Sistemas Baseados em Conhecimento), que representavam o conhecimento humano de um determinado domínio por meio da codificação explícita de regras para resolução de tarefas específicas. Esse conhecimento era extraído de especialistas por meio de entrevistas, nas quais se buscava identificar as regras utilizadas na tomada de decisão. No entanto, esse processo apresentava limitações, como subjetividade e

resistência dos especialistas, muitas vezes motivada pelo receio de que a transferência do conhecimento os tornasse dispensáveis (Faceli *et al*., 2021).

Com o avanço da capacidade computacional e a crescente disponibilidade de grandes volumes de dados os modelos analíticos baseados no Aprendizado de Máquina (AM) (*Machine Learning* - ML) tornaram-se mais viáveis e eficientes, permitindo a extração automática de padrões a partir dos dados, reduzindo a necessidade de regras programadas manualmente e viabilizando a execução de tarefas cognitivas complexas, como processamento de linguagem e detecção de padrões (Janiesch; Zschech; Heinrich, 2021).

O AM pode ser definido, portanto, como "o ramo da inteligência artificial que tem a capacidade de obter novos conhecimentos com a experiência e realizar atividades não definidas explicitamente em suas instruções programadas" (Rahman, 2022, p. 39).

Para isso, o AM utiliza heurísticas para modelar padrões nos dados disponíveis, geralmente estruturados em formato tabular, onde cada linha é um objeto (instância ou exemplo) e cada coluna corresponde a um atributo (variável). Os atributos podem ser preditivos, que descrevem as características dos objetos e são frequentemente chamados de variáveis explicativas, variáveis independentes, características, atributos (*features*), preditores ou covariáveis. Já o atributo-alvo representa a variável que o modelo busca prever com base nos atributos preditivos de um conjunto de dados (*dataset*). Também chamado de variável resposta ou variável dependente, ele pode ser classificado como categórico ou contínuo (Faceli *et al*., 2021; Izbicki; Santos, 2020).

Conforme aponta Schmidhuber (2015), os principais paradigmas de AM são o aprendizado supervisionado, o aprendizado não supervisionado e o aprendizado por reforço. No aprendizado supervisionado, cada instância nos dados de treinamento (conjunto de exemplos utilizados para ensinar o modelo) possui um atributo-alvo com um valor conhecido, chamado de rótulo (*label*). O rótulo serve como referência para que o modelo aprenda a relação entre os atributos preditivos e o atributo-alvo.

Na aprendizagem não supervisionada, muitas vezes chamada de descritiva, não se tem acesso a um conjunto de treinamento rotulado, o algoritmo tenta aprender padrões sem ter informações explícitas sobre as saídas desejadas. O termo descritivo refere-se à capacidade da aprendizagem não supervisionada de explorar e descrever a estrutura subjacente dos dados.

Na aprendizagem por reforço, o agente não recebe exemplos rotulados ou uma variável alvo explícita, como acontece na aprendizagem supervisionada. Aqui, um agente deve aprender como se comportar em um ambiente dinâmico através de interações do tipo "tentativa e erro".

Para isso ele recebe uma recompensa após cada ação, mas não sabe imediatamente qual ação é a melhor para alcançar o objetivo final. Ele precisa aprender, através da experiência, a navegar no ambiente para otimizar suas decisões. (Maxsuell, 2017).

O treinamento em AM é o processo de ajuste dos parâmetros de um modelo a partir de dados de treinamento, utilizando uma função objetivo – dependendo do contexto, também chamada de função de perda (*loss function*), função de custo (*cost function*), função de erro (*error function*), função de recompensa (*reward function*) ou função de aptidão (*fitness function*) – para minimizar erros ou maximizar o desempenho do modelo, permitindo que ele generalize padrões para novos dados. Esse processo frequentemente utiliza um algoritmo de otimização para encontrar um padrão nos dados e gerar previsões mais precisas (Wilmott, 2019).

Cada um dos três paradigmas de AM pode empregar diferentes regras de aprendizagem, dependendo do tipo de problema a ser resolvido. Conforme abordado por Haykin (2007), essas regras determinam como os algoritmos ajustam seus parâmetros para otimizar o aprendizado e melhorar o desempenho. Entre elas, incluem-se o aprendizado por correção de erro, busca minimizar iterativamente uma função de custo, ajustando os pesos do modelo em relação a um conjunto de dados; o aprendizado baseado em memória, que se beneficia das experiências anteriores para tomar decisões atuais; o aprendizado por árvores de decisão, que organiza a tomada de decisões em uma hierarquia de perguntas e respostas; e o aprendizagem por agrupamento, que visa identificar padrões naturais nos dados.

Além disso, existem abordagens como o aprendizado incremental, que se concentra em atualizar modelos continuamente à medida que novos dados são introduzidos, sem a necessidade de aprender do zero. Outras abordagens incluem a aprendizagem hebbiana, que ajusta conexões em redes neurais artificiais com base na ativação conjunta de neurônios ('neurônios que disparam juntos, conectam-se juntos'), e a aprendizagem competitiva, na qual diferentes neurônios competem para responder a estímulos, levando à especialização em padrões específicos. Ambas são comumente usadas em modelos artificiais e possuem fundamentos que se conectam a processos biológicos observados em sistemas neurais.

De modo geral, trata-se de um campo amplo e dinâmico, em que novos conceitos podem surgir à medida que as pesquisas vão avançando nessa área. A Figura 2 sintetiza os principais paradigmas de AM juntamente com os principais algoritmos encontrados na literatura.

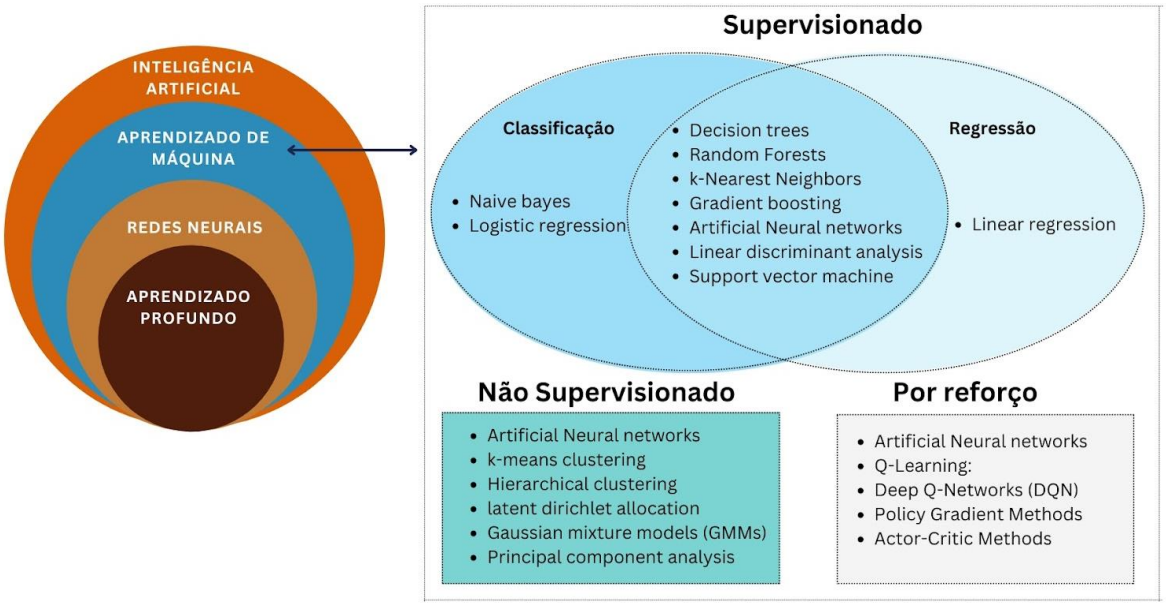

***Figura 2**: Abordagens do Aprendizado de Máquina e os principais algoritmos utilizados. Autor (2024).*

3.4.2 Processamento de Linguagem Natural (PLN)

A linguagem natural é aquela usada para a comunicação diária por seres humanos, em idiomas como inglês, espanhol, francês ou português. Ao contrário das linguagens artificiais, como as de programação e notações matemáticas, as linguagens naturais evoluíram ao longo de gerações e são difíceis de definir com regras explícitas, pois são moldadas por contextos culturais e sociais, tornando sua estrutura e uso mais flexíveis e dinâmicos.

Essa complexidade e dinamismo apresentam desafios para a modelagem computacional, o que nos leva ao campo do Processamento de Linguagem Natural (PLN). PLN é uma subárea da inteligência artificial que capacita máquinas a processar, compreender, analisar e gerar dados de linguagem natural, como texto e fala. Ele abrange uma ampla gama de tarefas, incluindo análise sintática, extração de informações, tradução automática, análise de sentimentos, classificação de textos e geração de linguagem natural. Conforme Khurana *et al*. (2023), o PLN pode ser classificado em duas partes, a Compreensão da Linguagem Natural (NLU), que envolve a tarefa de entender os diferentes níveis linguísticos (fonologia, morfologia, lexical, sintático, semântico, discursivo e pragmático), e a Geração da Linguagem Natural (NLG), que se concentra na produção de textos, como em resumos automáticos, respostas de chatbots ou geração de linguagem criativa.

O PLN cresceu significativamente nos últimos anos graças ao aumento no poder computacional, à disponibilidade de quantidades muito grandes de dados linguísticos, ao desenvolvimento de métodos de aprendizado de máquina (AM) altamente bem-sucedidos e a

uma compreensão muito mais rica da estrutura da linguagem humana e sua implantação em contextos sociais (Hirschberg; Manning, 2015).

Apesar desse crescimento, essa área ainda enfrenta alguns desafios e limitações. Dentre os desafios comuns, está a presença de parônimos, sinônimos e homônimos, que podem confundir os algoritmos devido a suas semelhanças fonéticas ou ortográficas, mas com diferentes significados. Além disso, palavras que transmitem sarcasmos e ironia; ambiguidade; frases informais, expressões idiomáticas e jargões complicam ainda mais a interpretação precisa; erros ortográficos e o uso incorreto de palavras também representam um desafio, pois podem alterar completamente o sentido pretendido de uma frase ou texto (Khurana *et al*., 2023).

Em termos de limitações no PLN, Bird *et al*. (2019) citam que, apesar de vários avanços, os sistemas de linguagem natural implantados para aplicativos do mundo real ainda não estão aptos a realizar o raciocínio de bom senso, ou seja, dificuldade em fazer julgamentos ou inferências que para os humanos parecem óbvios e intuitivos. Ou ainda, esses sistemas não conseguem aproveitar o conhecimento geral do mundo de forma eficaz, apresentando problemas para entender e aplicar informações gerais sobre como o mundo funciona.

Outra limitação está no fato da predominância de recursos e sistemas de PLN disponíveis apenas para idiomas com grandes volumes de dados textuais, a exemplo do inglês, o que limita o desenvolvimento de modelos eficazes para idiomas que dispõe de poucos dados disponíveis para treinamento (Srivastava *et al*., 2022). Essa desigualdade compromete a qualidade dos modelos, restringe sua generalização a outros contextos linguísticos e dificulta o acesso a tecnologias essenciais, como tradução automática e assistentes virtuais.

Nesse contexto, o pré-processamento textual desempenha um papel importante, pois ajuda a preparar os textos de forma mais consistente e padronizada, facilitando a análise subsequente sem comprometer a interpretabilidade do modelo. As principais etapas incluem padronização do texto (letras minúsculas, remoção de acentos e caracteres especiais), tokenização, remoção de *stopwords* e normalização. Além disso, técnicas como *stemming*, lematização e etiquetamento de partes do discurso também são usadas para melhorar a qualidade da análise (Denny; Spirling, 2018).

Aplicações de análise de dados e extração de conhecimento têm se utilizado largamente dos avanços no PLN. De modo que, para lidar efetivamente com o grande volume de dados textuais disponíveis, é preciso usar uma combinação de técnicas de representação textual e métodos de aprendizado de máquina.

3.4.3 Técnicas de Representação Textual

No campo do PLN, diversas técnicas de representação de palavras são usadas para transformar texto bruto em formatos que os computadores podem manipular. Naseem *et al*. (2021) exploraram as diferentes técnicas utilizadas, destacando os métodos clássicos — que englobam abordagens categóricas (*One-hot Encoding* e *Bow*) e ponderadas (TF e TF-IDF) — e os métodos mais recentes de *word embeddings*. Esses métodos serão abordados a seguir.

A forma mais simples de representar um texto é utilizando a técnica de *One-hot Encoding* que consiste em representar cada palavra de um vocabulário (um conjunto de palavras) por um vetor binário em que todos os elementos são 0, exceto um, que é 1 (indicando a posição única da palavra no vocabulário). Essa abordagem trata cada palavra como uma categoria distinta, sem qualquer relação semântica ou quantitativa entre elas.

A técnica *Bag-of-Words* (BoW) conta a frequência das palavras nos documentos, ao invés de simplesmente indicar presença ou ausência como no *One-hot Encoding*. Esse método converte sentenças em vetores de dimensão fixa, onde cada dimensão corresponde a uma palavra do vocabulário e é quantificada pela frequência dessa palavra no documento. Assim, cada índice no vetor representa a frequência de ocorrência da respectiva palavra no documento, sem manter qualquer informação sobre a posição original da palavra no texto (Goldberg, 2017).

Conforme Naseem *et al*. (2021), os métodos de representação ponderados incluem o TF-IDF, que combina a Frequência de Termo (*Term Frequency* - TF) e a Frequência inversa de documento (*Inverse Document Frequency* - IDF) para avaliar a importância das palavras em documentos com base tanto em sua frequência local quanto em sua raridade em todo o corpus.

O TF mede a frequência com que cada palavra aparece em um documento em relação ao total de palavras no documento, fornecendo uma medida da importância da palavra no contexto desse documento específico. O IDF é uma medida de quão comum ou raro um determinado termo é em todos os documentos. Essa técnica diminui a importância de palavras que aparecem em muitos documentos no corpus, considerando-as menos significativas para qualquer documento em particular. Essa técnica aumenta a importância de palavras que são raras. Embora muito utilizado em aplicações de PLN como classificação de texto, recuperação de informações e agrupamento de texto, o TF-IDF não capta bem a semântica das palavras (palavras com significados semelhantes têm pesos diferentes) e ignora a ordem das palavras e a estrutura gramatical. Para resolver esta limitação, as alternativas mais modernas de representação textual são os métodos de *Word Embeddings* (Naseem *et al*. 2021).

Um *word embedding* (ou apenas *embedding*) é a forma mais usual de representação aprendida para um texto e surgiu da necessidade de se expressar as características de similaridade entre as palavras, de forma que isso pudesse ser aproveitado no contexto das aplicações. É uma técnica que representa palavras, frases ou documentos como vetores densos e de baixa dimensão em um espaço vetorial contínuo, utilizada em métodos de redes neurais para a extração de características textuais, a qual mapeia as palavras de tal forma que palavras que estão mais próximas no espaço vetorial sejam semelhantes em significado (Paiva; Pereira, 2022).

Diferentemente das representações esparsas (muitos zeros) como *one-hot encoding*, os *embeddings* reduzem a dimensão do espaço de representação comparado ao vocabulário, reduzindo a complexidade computacional e o uso de memória.

Alguns métodos comuns para geração de *embeddings* incluem modelos estáticos (sem-contexto) como Word2Vec (Mikolov *et al*., 2013), GloVe (Pennington *et al*., 2014) e FastText (Bojanowski *et al*., 2017), que geram um único *embedding* fixo para cada palavra no vocabulário, independentemente do contexto em que ela aparece. Por outro lado, modelos dinâmicos (contextualizados), como o BERT (Devlin *et al*., 2018), geram *embeddings* que variam de acordo com o contexto textual em que a palavra é usada, resultando em representações mais ricas e adaptativas.

3.4.4 Redes Neurais Artificiais (RNAs)

As Redes Neurais Artificiais (RNAs) têm sido um paradigma interessante nas áreas de sistemas ligados à IA, uma vez que trazem aspectos inspirados “na estrutura e no funcionamento do sistema nervoso, com o objetivo de simular a capacidade de aprendizado do cérebro na aquisição de conhecimento” (Faceli *et al*., 2021, p. 101).

O estudo de McCulloch e Pitts (1943) introduziu o conceito de neurônio artificial, modelado matematicamente como uma unidade lógica com limiar (*Threshold Logic Unit, TLU*), capaz de executar funções lógicas básicas como *AND* e *OR*. Embora tenham demonstrado o potencial computacional da combinação de múltiplas TLUs, esses modelos iniciais não possuíam capacidade de aprendizado.

Posteriormente, Hebb (1949) contribuiu com estudos sobre aprendizado neural, sugerindo que conexões sinápticas se fortalecem quando os neurônios disparam juntos de maneira repetida. Embora não tenha proposto um modelo computacional treinável, essa ideia influenciou o desenvolvimento de mecanismos de aprendizado em RNA, como o aprendizado

hebbiano e técnicas de ajuste de pesos, especialmente modelos baseados em aprendizado não supervisionado.

Em 1958, Rosenblatt introduziu o Perceptron, o primeiro modelo treinável de RNA, baseado no aprendizado supervisionado por correção de erros. O Perceptron de Rosenblatt (1958) era um modelo de camada única com apenas um neurônio. Minsky e Papert (1969) demonstraram matematicamente que o modelo de camada única tinha a limitação de não ser capaz de resolver problemas que não fossem linearmente separáveis, como a implementação da porta lógica XOR (Rosenblatt, 1958). Essa limitação foi superada com a criação das redes neurais de múltiplas camadas, também conhecidas como Perceptron de Múltiplas Camadas (*Multilayer Perceptron,* MLP). As MLPs são um tipo de rede neural *feedforward*, compostas por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, nas quais o sinal se propaga em uma única direção, da entrada para a saída, sem conexões recorrentes (Haykin, 2007).

Uma RNA possui dois elementos: uma arquitetura (topologia da rede) e um algoritmo de aprendizagem (método pelo qual os pesos sinápticos são ajustados). A arquitetura define a organização da rede, incluindo o número de camadas, a quantidade de neurônios por camada e a forma como estão conectados. Já o algoritmo de aprendizado determina como os pesos sinápticos são ajustados com base nos dados de entrada e saída, permitindo que a rede aprenda padrões a partir da experiência e armazene esse conhecimento nas conexões ponderadas entre os neurônios (Faceli, 2021).

Outro elemento importante a considerar com relação ao aprendizado é a capacidade de generalização de uma RNA. O aprendizado envolve o desenvolvimento de um conhecimento experimental: uma RNA aprende a partir de uma série de exemplos ou amostras provenientes de seu ambiente. Para que uma RNA desempenhe bem sua tarefa, é necessário que seja robusta o suficiente para poder inferir sobre uma série de exemplos semelhantes ou distintos entre si ou mesmo sobre novos exemplos adicionados ao contexto que não tenham feito parte do desenvolvimento do conhecimento experimental.

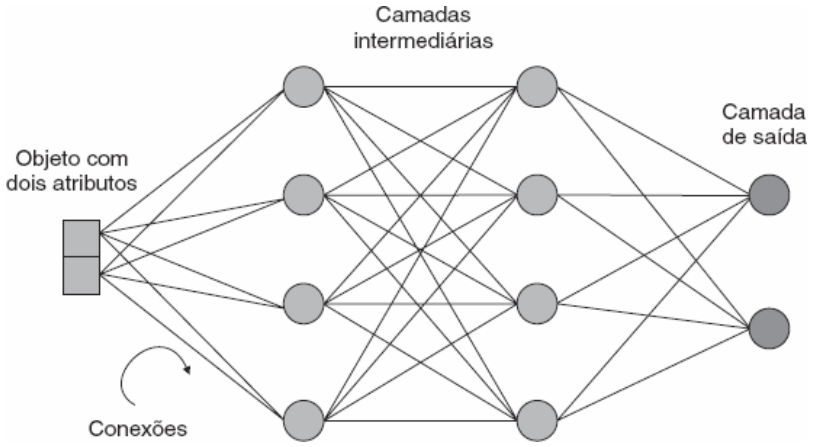

***Figura 3**: Arquitetura de RNA múltiplas camadas. Fonte: Faceli, 2021.*

A Figura 3 ilustra uma arquitetura com duas camadas ocultas. A camada de entrada é responsável apenas pela recepção e distribuição dos dados de entrada para as camadas subsequentes (propagação), sem participar do aprendizado. Os sinais de entrada estão conectados a cada neurônio das camadas ocultas que processam esses sinais multiplicando os valores de saída da camada anterior por um vetor de pesos (W) e adicionando um *bias* (b). Esses valores são então passados por uma função de ativação, gerando um valor de saída que contribui para o aprendizado da rede. Este ajuste de pesos e *bias* é realizado com base nos erros calculados durante o treinamento (retropropagação). A camada de saída utiliza esses dados processados para inferências, determinando valores em problemas de regressão ou categorias em problemas de classificação. O *bias* ajusta o limiar de ativação da função, melhorando a capacidade de resposta da rede a diferentes intervalos de entrada, sendo ajustado junto aos pesos para otimizar os resultados da aprendizagem (Haykin, 2007).

O aprendizado de uma RNA ocorre por meio do ajuste iterativo dos pesos sinápticos, de forma a minimizar uma função de custo. O processo de aprendizado inicia-se com a inicialização aleatória dos pesos dos neurônios. Durante o treinamento supervisionado, por exemplo, a rede produz um resultado para cada exemplo de entrada, e o erro de cada previsão é calculado, geralmente por meio do gradiente do erro. Este erro é então propagado de volta pela rede (*backpropagation*), ajustando os pesos dos neurônios para minimizar a função de custo, melhorando assim a precisão das previsões da rede.

As RNAs evoluíram consideravelmente desde suas primeiras concepções, resultando no desenvolvimento de outras arquiteturas especializadas. Além dessas Redes Neurais Feedforward (FNNs), que processam informações de forma unidirecional, surgiram outras arquiteturas, como as Redes Neurais Convolucionais (CNNs - *Convolutional Neural Networks*),

projetadas para lidar com dados estruturados espacialmente, como imagens e vídeos (LeCun *et al*., 1989). Essas redes incorporam camadas de convolução, que identificam padrões e características em diferentes níveis de abstração, e camadas de pooling, responsáveis por reduzir a dimensionalidade e otimizar o processamento.

Para dados sequenciais, as Redes Neurais Recorrentes (RNNs - *Recurrent Neural Networks*) introduzem conexões temporais que permitem capturar dependências entre elementos de uma sequência. No entanto, RNNs convencionais sofrem com o problema de desvanecimento e explosão do gradiente (Pascanu; Mikolov; Bengio, 2012), dificultando a retenção de infrmações de longo prazo. Outras arquiteturas foram desenvolvidas para reduzir esse problema, como as *Long Short-Term Memory* (LSTMs) (Hochreiter; Schmidhuber, 1997) e *Gated Recurrent Units* (GRUs) (Cho *et al*.,2014) que incorporam mecanismos de controle de fluxo de informação.

3.4.5 Modelos baseados em Atenção

Antes da ascensão dos modelos baseados em atenção, a literatura aponta diferentes tecnologias e abordagens no campo do processamento de linguagem natural que foram usadas para resolver tarefas relacionadas com o processamento de texto e linguagem. Essas tecnologias e abordagens anteriores incluem modelos baseados em regras, métodos estatísticos, máquinas de vetores de suporte (*Support Vector Machine* - SVM) e outros algoritmos de aprendizado de máquina e redes neurais recorrentes (*Recurrent Neural Network* - RNN) (Hajikhani; Suominen, 2022).

Apesar dos avanços trazidos por essas tecnologias, a busca por modelos mais eficazes e adaptáveis levou ao desenvolvimento da arquitetura codificador-decodificador (Cho *et al.*, 2014), que se destacou em tarefas de processamento de linguagem natural, como a tradução automática explorada por Sutskever *et al*. (2014).

Nessa arquitetura, o codificador processa cada item na sequência de entrada, gerando uma representação intermediária ou vetor de contexto (*context vector*) que contém informações sobre toda a sequência de entrada. Depois de processar toda a sequência de entrada, o codificador envia esse vetor para o decodificador, que, por sua vez, usa essa informação para produzir a saída desejada, seja uma tradução, uma classificação, uma resposta gerada automaticamente, entre outras possibilidades, uma vez que essa arquitetura pode ser aplicada a uma variedade de tarefas de PLN.

No entanto, esses modelos tradicionais, como RNNs e LSTMs (*Long Short-Term Memory*), enfrentam dificuldades em lidar com dependências de longo prazo em sequências extensas, além de incorrer em sobrecarga de cálculo para realizar a operação de decodificação e treinamento demorado devido ao problema de desvanecimento e explosão do gradiente. Para resolver tais problemas, foi desenvolvido o mecanismo de atenção (Bahdanau *et al*., 2014), inicialmente utilizado em modelos de tradução automática com RNNs (Luong *et al*., 2015).

Os mecanismos de atenção são métodos computacionais que imitam a atenção humana, permitindo que algoritmos focalizem partes específicas dos dados de entrada. No PLN, a atenção neural detecta as correlações existentes nos componentes do texto, mapeando a estrutura da linguagem e suas relações de contexto e ordenação.

Hu (2020), apresenta em sua revisão diferentes variantes do mecanismo de atenção propostas para lidar com uma variedade de tarefas em PLN, entre elas a autoatenção. O mecanismo de autoatenção (*self-attention*) é usado integralmente na arquitetura de rede neural *Transformers*, apresentada em 2017 no trabalho de Vaswani *et al*. (2017). Esse mecanismo é usado para lidar com sequências de *tokens* (palavras), permitindo que o modelo capture relações complexas e de longo alcance entre os elementos da sequência de entrada.

No mecanismo de autoatenção, cada *token* na sequência de entrada é representado por seus respectivos vetores: vetor de consulta (Q), vetor de chave (K) e vetor de valor (V), eles são derivados do vetor de entrada por meio de transformações lineares separadas. Em seguida, é calculada uma matriz de atenção não normalizada através dos produtos escalares entre os vetores Q e K, que representa a similaridade entre todos os pares de *tokens* na sequência.

Estas matrizes são normalizadas utilizando a função *Softmax* para obter os pesos de atenção que representam a importância relativa de cada *token* em relação aos outros *tokens* na sequência de entrada. Esses pesos de atenção são multiplicados pelos vetores de valor (V), resultando em um vetor de contexto, que é uma combinação ponderada dos vetores de entrada. A autoatenção é calculada paralelamente para todos os *tokens* na sequência de entrada, permitindo ao modelo obter uma representação mais abrangente da sequência de *tokens* em tarefas de PLN.

Transformers

Transformer é uma abordagem recente e muito utilizada no aprendizado sequência para sequência (*seq2seq*), baseada em mecanismos de autoatenção que permitem aprender dependências de longo alcance entre sequências de entrada e saída.

Essa arquitetura se destaca por sua alta capacidade de paralelização. Em termos de aprendizado de máquina e redes neurais, isso se traduz em realizar cálculos em várias partes dos dados ou da arquitetura do modelo simultaneamente para economizar tempo de processamento. Durante o treinamento ou inferência, as operações matemáticas envolvidas podem ser divididas e calculadas em paralelo em unidades de processamento, como GPUs (*Graphics Processing Units*) ou TPUs (*Tensor Processing Units*). Isso acelera o tempo necessário para treinar e usar o modelo, superando arquiteturas recorrentes e convolucionais, resultando em tempos de treinamento significativamente reduzidos (Vaswani *et al*., 2017).

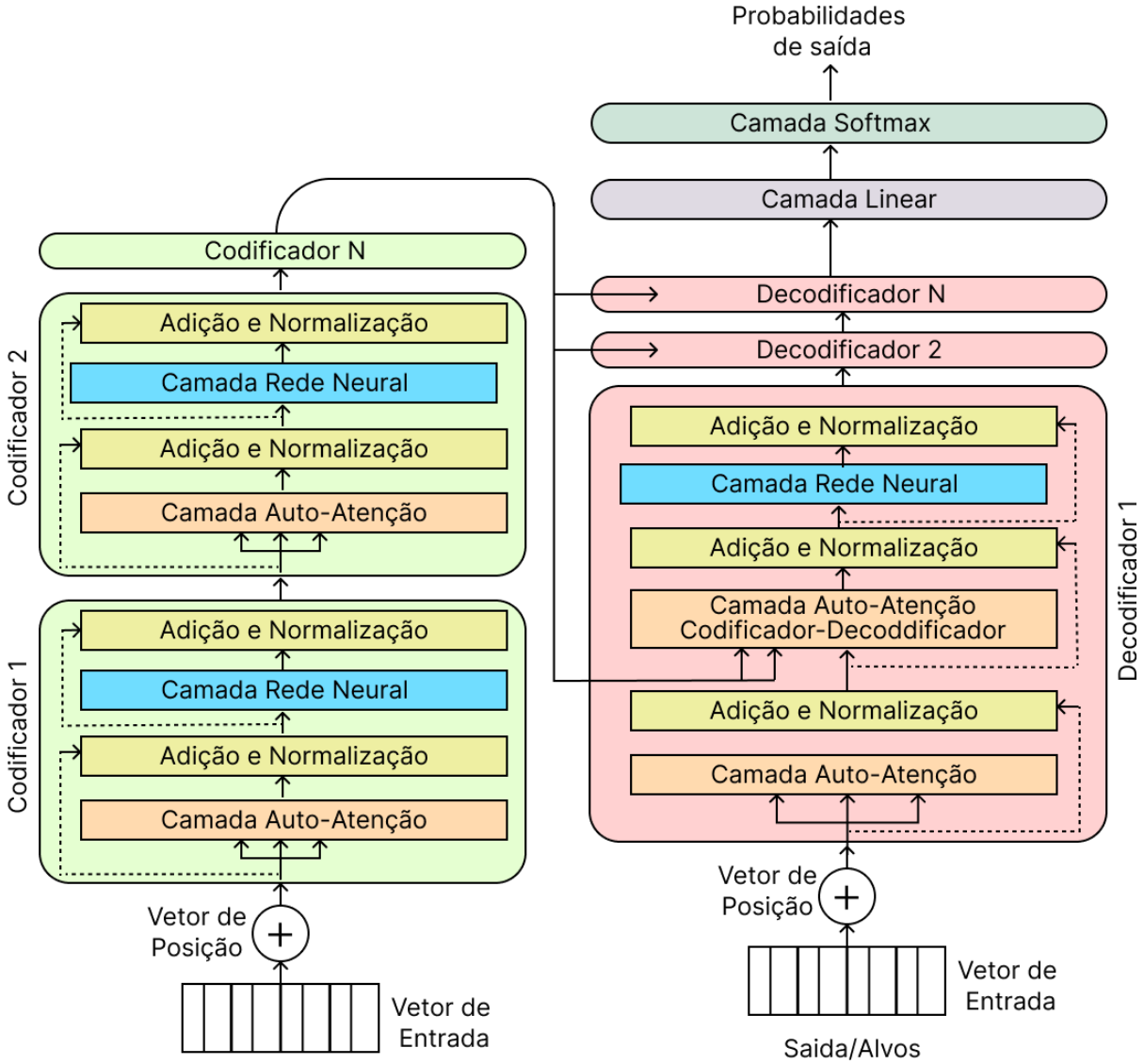

***Figura 4**: Modelo da Arquitetura Transformer. Fonte: Autor (2024), adaptado de Vaswani et al., 2017.*

A figura 4 ilustra a arquitetura geral do *Transformer*. Destacam-se dois componentes principais: As pilhas de codificadores - Nx camadas de codificador idênticas (no artigo original

publicado, Nx = 6). Cada camada codificadora tem duas subcamadas, sendo que a primeira delas é um mecanismo de autoatenção multifoco e a segunda é uma rede neural *feedforward* totalmente conectada (*Fully Connected Feedforward Network*).

O segundo componente são as pilhas de decodificadores - Nx camadas de decodificadores idênticas (no artigo original publicado, Nx = 6). Cada camada decodificadora possui três subcamadas. A primeira subcamada é o mecanismo de autoatenção multifoco (*Masked Multi-Head Self-Attention*), semelhante à do codificador, mas aplicada à sequência de entrada do decodificador. Aqui, há um mascaramento causal aplicado para garantir que a posição *i* não possa atender à posição *i+1* e subsequentes, garantindo que cada previsão dependa somente das saídas anteriores.

A segunda subcamada do decodificador, é um mecanismo de atenção cruzada (*Cross-Attention*), pois opera entre duas sequências diferentes. Nesta segunda subcamada cada posição em uma sequência (a sequência de saída gerada até o momento atual) atende a todas as

posições na saída do codificador. Isso permite que o decodificador atenda a diferentes partes da representação codificada da entrada. Terceira subcamada, semelhante à do codificador, é uma rede neural *Feedforward* totalmente conectada.

Arquiteturas baseadas em codificador-decodificador *(encoder-decoder*) do *Transformer* é usado, por exemplo, para tarefas como tradução de idiomas, presente nos sistemas como o Google Translator e o M2M-100 do Facebook (Fan *et al*., 2020). Um exemplo de modelo baseado nessa arquitetura é o T5 (*Text-to-Text Transfer Transformer*), desenvolvido pelo Google (Raffel *et al*., 2019). No T5 todas as tarefas são formuladas como problemas de transformação de texto para texto. O codificador recebe a sentença de entrada e produz uma representação vetorial, que é alimentada no decodificador para gerar a sentença de saída. O decodificador usa autoatenção para processar a sequência de entrada e atenção cruzada, onde o mecanismo de atenção é aplicado tanto à saída quanto à entrada do decodificador.

Já a parte do codificador do *Transformer* é usada, por exemplo, para tarefas como classificação de texto, onde o modelo deve classificar um trecho de texto em uma das diversas categorias predefinidas, como proposto neste trabalho.

Modelos da família GPT (*Generative Pre-trained Transformer*), usado para geração de texto, é composto apenas por camadas de decodificador, enquanto outros, como o BERT que será abordado a seguir, utiliza apenas camadas codificadoras do *Transformer*.

BERT

O *Bidirectional Encoder Representations from Transformers* (BERT), desenvolvido pelo Google (Devlin *et al*., 2018), consiste em uma pilha de camadas codificadoras baseadas na arquitetura do *Transformer* que incorpora objetivos específicos e uma abordagem com múltiplas cabeças de autoatenção. Ele faz parte da família de modelos fundacionais (Bommasani *et al*., 2022), e que podem ser adaptados a uma ampla gama de tarefas. Está disponível em duas variantes principais: o *BERT-Base* (12 camadas) e o *BERT-Large* (24 camadas). Embora o *BERT-Large* seja capaz de capturar relações mais complexas entre as palavras devido ao seu maior número de camadas, também requer mais recursos computacionais para treinamento e inferência.

Em comparação com modelos autorregressivos como o GPT, que geram previsões principalmente com base no contexto anterior à palavra atual, e modelos que consideram apenas o contexto à direita de uma palavra (*right-to-left models*), o BERT utiliza uma abordagem bidirecional, ele é treinado para prever palavras em uma sentença dada toda sua informação contextual.

O BERT é pré-treinado (*pre-training*) em tarefas de aprendizado não supervisionado utilizando grandes quantidades de dados (aproximadamente 3.3 bilhões de palavras extraídas de livros e artigos online) de texto não rotulados, permitindo que ele aprenda representações de linguagem geral. O pré-treinamento é a abordagem dominante para transferência de aprendizagem, onde um modelo é treinado em uma tarefa substituta (muitas vezes apenas como um meio para um fim) e depois adaptado à tarefa de interesse posterior por meio de ajuste fino (Bommasani *et al*., 2022).

Durante o pré-treinamento, o BERT é submetido a duas tarefas não supervisionadas, a Previsão de Palavras Mascaradas (*Masked Language Model* - MLM) e a Previsão de Sentenças Seguinte (*Next Sentence Prediction* - NSP). Na tarefa MLM, um percentual de palavras é aleatoriamente mascarado (substituídas por um *token* [MASK]), e o modelo tenta prever essas palavras com base no contexto das palavras vizinhas. A tarefa NSP fornece ao modelo pares de sentenças reais e falsas, criadas combinando partes de diferentes sentenças a fim de treiná-lo para prever se a segunda sentença é, de fato, a sequência seguinte da primeira e diferenciá-las corretamente. Essas tarefas permitem ao modelo compreender a relação entre diferentes partes do texto e capturar informações de contexto mais amplo (Islek; Oguducu, 2020).

No ajuste fino (*fine-tuning*), o modelo é treinado em tarefas específicas com dados rotulados, ajustando os parâmetros e pesos da rede neural para maximizar o desempenho na tarefa em questão (Rogers *et al*., 2020).

Conforme destacou Kaliyar *et al*. (2021), nesta etapa há três características importantes a serem consideradas. Primeiro, é o pré-processamento de texto longo, pois o *BERT-base*, por exemplo, modelo mais comumente utilizado, possui um limite máximo de comprimento de sequência de 512 *tokens*. Contudo, existem variantes da arquitetura BERT com capacidade estendida para lidar com textos mais longos, como o Longformer (Beltaby *et al*., 2020), que pode processar sequências de até 4096 *tokens*. Dessa forma, ao lidar com textos longos, é necessário garantir que eles se encaixem nesse limite. Segundo, é importante considerar a seleção de camadas, pois o modelo base do BERT é composto por uma camada de incorporação e 12 camadas de codificador; Terceiro, o problema de sobre-ajuste, que ocorre quando um modelo se ajusta muito bem aos dados de treinamento, mas tem dificuldade em generalizar para novos dados.

De fato, o BERT tem sido aplicado em uma variedade de outras tarefas de PLN (Qiu *et al*., 2021), como, por exemplo, classificação de documentos (Adhikari, 2019), sistemas de recomendação (Ray *et al*., 2021), reconhecimento de entidades nomeadas (Virtanen *et al*., 2019), análise de sentimento (Zhang *et al*., 2020) e resposta a perguntas (Yang; Choi, 2019). O sucesso desta arquitetura está diretamente ligado à sua capacidade de compreender as nuances e sutilezas da linguagem natural, capturando o contexto e as relações entre as palavras de forma eficaz.

Quando foi publicado, o BERT revolucionou o campo do PLN. Desde então, outras variantes foram desenvolvidas para otimizar o modelo e adaptá-lo a diferentes fins. Entre elas, destacam-se o DistilBERT, uma versão compacta do modelo BERT que foi reduzida em tamanho, mantendo 97% de sua capacidade de compreensão do idioma e sendo 60% mais rápido (Sanh *et al*. 2020); o RoBERTa (Liu *et al*., 2019) treinado em mais de 30B palavras, obteve melhores resultados que o BERT em tarefas de perguntas e respostas e compreensão da linguagem; o XLNet (Yang *et al*., 2020), treinado em 33B de palavras, apresentou melhor desempenho que o BERT em 20 tarefas; o GPT-3 (Brown *et al*., 2020) treinado em um corpus com aproximadamente 400B de palavras, representa a terceira iteração da série de modelos GPT que ganhou popularidade, especialmente, através da aplicação ChatGPT, destacando-se por sua capacidade de conversação e geração de texto.

Adicionalmente, modelos especializados como BioBERT (Lee *et al*., 2020), SciBERT (Beltagy *et al*., 2019), BERTabaporu (Da Costa *et al*., 2023) e BERTweet (Nguyen *et al*., 2020) foram adaptados para tarefas ou domínios específicos. Essas variantes mantêm a arquitetura *Transformer* subjacente, proporcionando avanços significativos em diversas tarefas de PLN.

Neste estudo, foram empregados três modelos de linguagem pré-treinados em corpora do português do Brasil, cada um caracterizado por sua arquitetura e adaptação específicas para lidar com tarefas de PLN nesse contexto linguístico. Eles são apresentados a seguir:

Albertina PTBR 100M (Rodrigues *et al*., 2023): este modelo foi treinado em uma seleção de 3.7 bilhões de *tokens* do conjunto de dados OSCAR, que inclui documentos em mais de cem idiomas, incluindo o português. Ele é uma derivação da arquitetura DeBERTa (He *et al*., 2020). Esse modelo foi projetado para otimizar a compreensão e geração de texto em português brasileiro, incorporando características específicas do idioma.

BERTimbau (Souza *et al*., 2020): este modelo alcançou o estado da arte em reconhecimento de entidade nomeada (NER), similaridade textual de sentença (STS) e reconhecimento de implicação textual (RTE) em português brasileiro. Neste estudo foi utilizada a variante *BERTimbau Base* treinada no BrWaC (*Brazilian Portuguese Web as Corpus*).

DeBERTinha (Campiotti *et al*., 2023): uma adaptação do modelo DeBERTaV3 XSmall pré-treinado em inglês para tarefas de PNL em português brasileiro, com 40 milhões de parâmetros. Além do BrWac, o DeBERTinha também utilizou o conjunto de dados Carolina.

3.4.6 Métricas de Avaliação

Avaliar a qualidade de um modelo requer o uso de métricas quantitativas relevantes para o problema em questão. Neste estudo, foram empregadas a precisão, *recall* e *F1-Score* como medidas de avaliação de desempenho.

De acordo com Rezaeenour *et al*. (2022) a métrica de precisão (*precision*) em avaliação de classificadores visa determinar a proporção de amostras corretamente classificadas como positivas. Ela é calculada pela divisão do número de verdadeiros positivos pelo total de positivos identificados, conforme Equação 1.

$$precision = \frac{verdadeiros\ positivos}{verdadeiros\ positivos + falsos\ positivos} \quad (1)$$

A métrica de recall opera sobre a mesma linha da métrica de precisão, mas utiliza as amostras falsas negativas. Sua fórmula utiliza a razão da quantidade de verdadeiros positivos pela soma de verdadeiros positivos e falsos negativos, conforme Equação 2. Essa métrica mede a completude do classificador, em que um valor baixo indica um alto índice de falsos positivos.

$$recall = \frac{Verdadeiros\ Positivos}{Verdadeiros\ Positivos\ /\ Falsos\ Positivos} \quad (2)$$

Já a *F1-score* é a média harmônica da precisão e *recall*, fornecendo uma métrica balanceada que considera tanto falsos positivos quanto falsos negativos, conforme Equação 3. É útil quando há um desequilíbrio entre os rótulos.

$$F1\ score = 2 . \frac{Precision\ .recall}{Precision + recall} \quad (3)$$

3.5 Integração da Agenda 2030 e a Inteligência Artificial

Mais de três décadas após o Relatório Brundtland, o mundo ainda luta com a necessidade de transformar as atividades humanas para garantir a sobrevivência a longo prazo. Os desafios são complexos e exigem toda a capacidade e expertise tecnológica para desenvolver soluções rápidas e sustentáveis. Entre as possibilidades mais promissoras está a IA, que permite que máquinas aprendam com experiências, adaptem-se a novas situações e realizem atividades similares às humanas (Russel; Norvig, 2022).

A integração dos ODS com a IA representa uma convergência favorável para enfrentar desafios globais e promover o desenvolvimento sustentável, uma vez que, tanto a IA, como outras tecnologias digitais emergentes, como a internet das coisas (*internet of things* - IoT), *big data*, *blockchain*, realidade virtual e aumentada etc., estão sendo aplicadas em praticamente todas as áreas da sociedade, economia e meio ambiente.

Em sua revisão, Nishant *et al*. (2020) identificaram que as pesquisas sobre IA aplicada à sustentabilidade atravessam várias disciplinas e focam em desafios ambientais, onde áreas como água, energia, transporte e biodiversidade tem se mostrado promissoras. Por exemplo, pesquisas em sistemas hídricos, têm utilizado técnicas de aprendizado de máquina, como regressão linear, redes neurais artificiais (ANN) e máquinas de vetores de suporte (SVM), entre outras, para avaliar e prever o esgotamento de água subterrânea, analisar o fluxo de córregos e

examinar parâmetros de qualidade da água (Demirci *et al*., 2019; Khan *et al*., 2023). Estes estudos contribuem diretamente para o alcance dos ODS relacionados à gestão sustentável de recursos hídricos.

A Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), através do seu Programa Hidrológico Intergovernamental (IHP), também utiliza uma aplicação de IA que está em desenvolvimento desde 2005, em parceria com uma equipe da Universidade da Califórnia, Irvine (EUA). O aplicativo UNESCO G-WADI Geoserver[6] usa um algoritmo de rede neural artificial para estimar a precipitação em tempo real em todo o mundo e está sendo usado para informar o planejamento de emergência e o gerenciamento de riscos hidrológicos, como inundações, secas e tempestades. Um exemplo de aplicabilidade, foi o rastreamento do Super Tufão Haiyan conforme ele se aproximava das Filipinas em 2013 (conhecido como Super Tufão Yolanda, foi um dos ciclones tropicais mais poderosos já registrados). Além disso, serviços Hidrológicos de Seca da Namíbia (um país no sudoeste da África) usam o geoserver para preparar boletins diários com informações atualizadas sobre condições de inundação e seca para comunidades locais. Essa aplicação faz parte dos esforços da UNESCO para melhorar a gestão de recursos hídricos em regiões áridas e semi-áridas globalmente, contribuindo para o gerenciamento ambiental sustentável e preparação para desastres explorando a capacidade da IA (UNESCO, 2019).

Além disso, a integração da IA com os ODS tem sido vista também em aplicações que estão contribuindo significativamente na área da Saúde. As capacidades preditivas da IA estão sendo usadas para prever surtos de doenças como Dengue, Zika e Chikungunya em vários países. A *startup* Inteligência Artificial em Epidemiologia Médica (AIME, *Artificial Intelligence in Medical Epidemiology*) combina epidemiologia, análise de *big data* e IA para prever o momento e a localização de surtos de doenças infecciosas em tempo real. Especificamente no Brasil, a AIME estabeleceu uma parceria com a ONG Viva Rio para implementar um projeto piloto focado no combate e prevenção de doenças como Zika e Dengue . Atualmente operando com uma taxa de precisão de 87%, a ferramenta já foi implantada em projetos piloto não apenas no Brasil, mas também na Malásia e Filipinas (AIME, 2024; UNESCO, 2019; SDG NEWS, 2024).

Na pesquisa energética, além das redes neurais artificiais, tem-se aplicado sistemas especialistas, reconhecimento de padrões e modelos de lógica fuzzy em diversas áreas do setor.

[6] Disponível em: https://gwadi.org. Acesso em 28 ago 2024.

Essas tecnologias têm sido utilizadas para previsão de demanda de energia eólica e solar, otimizar o uso de energia em edifícios, melhorar a produção e distribuição de energia e aprimorar operações e manutenção (Merizalde *et al*., 2019; Sharifzadeh *et al*., 2019).

YE *et al*. (2020) mostram que a IA está sendo amplamente adotada para melhorar o controle da poluição ambiental, sendo empregada no tratamento de águas residuais, na redução da poluição do ar e no manejo de resíduos sólidos, desempenhando um papel importante na modelagem, otimização, previsão e controle desses processos.

Embora a integração da IA com os ODS ofereça oportunidades e avanços, ela também traz alguns desafios que precisam ser gerenciados. Um estudo realizado por Vinuesa *et al*. (2020) explora os impactos positivos e negativos dessa interseção, destacando a importância de enfrentar alguns desafios na área de IA. Um desses desafios é a interpretabilidade dos algoritmos, uma vez que é difícil entender como os algoritmos de IA tomam decisões, especialmente em casos complexos, podendo criar barreiras à adoção e à confiança em sistemas automatizados.

No entanto, os autores sugerem uma perspectiva otimista: a falta de interpretabilidade pode ser encarada como uma oportunidade para identificar e corrigir preconceitos e políticas injustas que possam estar sendo refletidos nos resultados desses algoritmos. Isso destaca a importância de não apenas avançar no desenvolvimento técnico da IA, mas também a consideração ética e a responsabilidade em garantir que a tecnologia seja aplicada de maneira transparente para promover os ODS.

A aplicação prática da IA nos ODS também enfrenta desafios na área de Processamento de Linguagem Natural (PLN), uma subárea da IA que envolve a engenharia de modelos e processos computacionais para resolver problemas práticos voltados a compreender as linguagens humanas (Otter *et al*., 2021), especialmente quando se lida com diferentes idiomas. Cada idioma apresenta suas próprias complexidades na coleta de documentos e no pré-processamento do corpus de texto. Isso implica que a abordagem e as técnicas utilizadas para coletar e processar documentos em um idioma podem não ser diretamente aplicáveis ou eficientes em outro idioma devido a diferenças nas estruturas linguísticas, disponibilidade de recursos e características específicas do idioma (Matsui *et al*., 2022).

Ademais, o uso de PLN e outras tecnologias de IA avançaram significativamente em sua aplicação aos ODS, permitindo análise de dados, previsão e tomada de decisões. As técnicas mais avançadas de PLN, por exemplo, são capazes de processar dados não estruturados, como texto ou linguagem humana, onde a informação pode estar dispersa, desorganizada ou em

formatos não convencionais. Os algoritmos de IA facilitam a modelagem preditiva e sistemas inteligentes, que são capazes de aprender com dados e fazer previsões ou tomar decisões com base nesse aprendizado. Essas tecnologias contribuem para monitorar indicadores e superar desafios relacionados ao desenvolvimento sustentável.

4 METODOLOGIA

Esta seção descreve os procedimentos metodológicos adotados nesta pesquisa. Vergara (2004) define metodologia como o estudo da instrumentação para a captação ou manipulação da realidade, envolvendo caminhos e procedimentos para alcançar determinados fins. Esta pesquisa é classificada como descritiva e exploratória, com o objetivo de investigar como universidades baianas contribuem para a Agenda 2030, correlacionando dados sem manipulá-los (Gil, 2009; Castro, 2006). A abordagem adotada é quantitativa, de natureza aplicada, utilizando-se de pesquisa documental e bibliográfica como estratégias principais para coleta e análise dos dados.

4.1 Construção do modelo ODSBAHIA-PTBR

O modelo ODSBAHIA-PTBR[7] foi desenvolvido para avaliar de forma autônoma e objetiva como descrições textuais em português se alinham aos ODS. Diferente de outras abordagens, o modelo incorpora, além dos 17 ODS oficiais da ONU, três novos objetivos propostos por Cabral e Gehre (2020) — Igualdade Racial (ODS 18), Arte, Cultura e Comunicação (ODS 19) e Direitos dos Povos Originários e Comunidades Tradicionais (ODS 20) —, visando ampliar a representatividade de grupos historicamente invisibilizados.

A estrutura analítica de construção do modelo ODSBAHIA-PTBR é apresentada na Figura 5. Primeiramente, foi construído e validado um corpus multirrótulo em português, correspondente aos 17 ODS da ONU e os três novos ODS propostos. Em seguida, foram avaliados e comparados os modelos pré-treinados BERTimbau, Albertina e DeBERTinha para identificar o mais adequado ao corpus. O modelo selecionado foi então finamente ajustado, dando origem ao ODSBAHIA-PTBR, que foi avaliado em relação a outros modelos de classificação de textos, com base em métricas como precisão, *recall* e *F1-score*.

[7] Santos, Ê. S., & Moraes, L. E. ODSBAHIA-PTBR: A Natural Language Processing Model to Support Sustainable Development Goals. Revista De Gestão Social E Ambiental, 18(12), e010230, 2024. https://doi.org/10.24857/rgsa.v18n12-039.

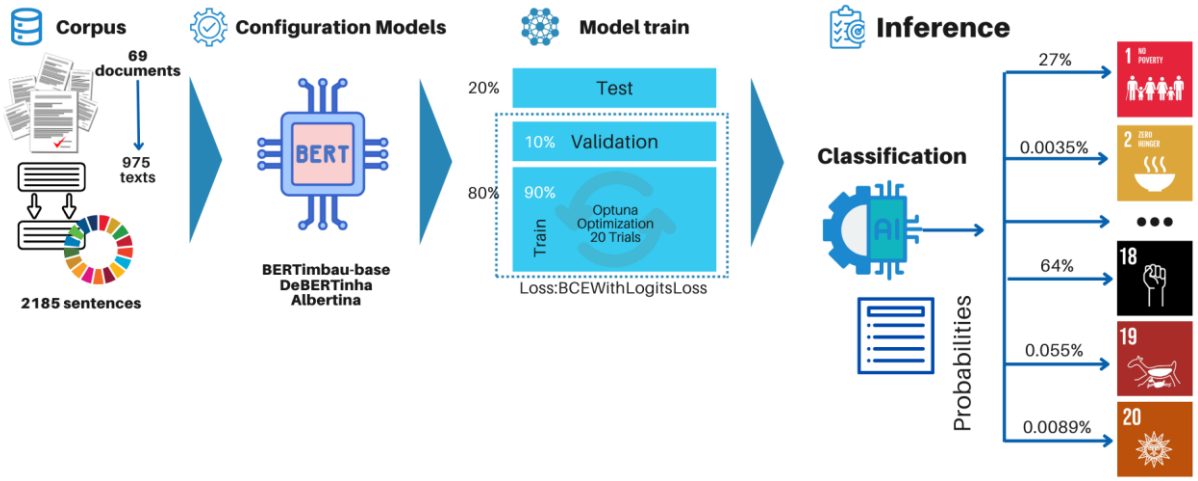

***Figura 5**: Estrutura analítica da metodologia do modelo ODSBahia-PTBR. Autor (2024)*

4.1.1 CONSTRUÇÃO E VALIDAÇÃO DO CORPUS DE TREINAMENTO

Não foi encontrado um corpus em português rotulado com os ODS; portanto, foi criado um corpus específico para esta tarefa a partir de trechos de texto em português, selecionados manualmente, principalmente dos Relatórios Luz de 2017 a 2023.

O Relatório Luz é produzido pelo Grupo de Trabalho da Sociedade Civil para a Agenda 2030 no Brasil (GT Agenda 2030), composto por diversas organizações da sociedade civil brasileira dedicadas ao monitoramento e avaliação da implementação dos ODS no país, com consulta a especialistas sobre os diversos temas contemplados na Agenda 2030. Esses relatórios têm como objetivo fornecer uma análise independente e crítica do progresso do Brasil em relação aos ODS, abrangendo diversas dimensões, tais como indicadores, políticas públicas, avanços e desafios. Os trechos de texto selecionados contêm informações contextualizadas sobre os ODS associados a eles. Adicionalmente, foram coletados trechos de notícias e artigos científicos.

Ao todo, foram analisados 69 documentos, dos quais foram extraídos 975 trechos contendo entre 3 e 8 frases, resultando em 2185 sentenças. Esses documentos foram escolhidos por sua relevância e abrangência no contexto das políticas socioambientais e sua relação com os ODS. Em média, cada trecho contém aproximadamente 70 palavras. A maioria desses trechos, 614 ao todo, foi extraída do Relatório Luz. O critério de seleção baseou-se na clareza com que o contexto de cada trecho indicava os ODS relacionados. Nos casos em que o conteúdo abordava múltiplos temas, foram atribuídos ODS adicionais para capturar a complexidade das questões abordadas.

Os ODS foram representados em um vetor de 20 dimensões, onde cada posição indica a presença (1) ou ausência (0) de um ODS específico no trecho analisado (Tabela 1). Nesse caso, um vetor [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0] indica que o texto está relacionado aos ODS 2, 13, 15. A atribuição dos rótulos foi validada através de uma análise dos contextos apresentados.

Texto_PT	One_hot_labels
Sem esforços efetivos dos três poderes da República para reverter a sub-representação política de mulheres, negros, indígenas, LGBTQIP+, pessoas com deficiência e demais grupos sociais nos processos de tomada de decisão. As mulheres seguem ocupando apenas 16% das cadeiras da Câmara dos Deputados, enquanto são 51,8% da população; negros e negras são 24,4%35 dos parlamentares federais e 56,2% na composição populacional; e apenas uma deputada representa as 256 nações indígenas ainda existentes no país.	[0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1]
A diminuição de estrutura do Instituto Nacional de Colonização e Reforma Agrária (Incra) e um modelo econômico que favorece o agronegócio exportador de commodities afetam em particular os povos e comunidades tradicionais (indígenas, quilombolas, entre outros), inclusive disponibilizando ao mercado terras destinadas à redistribuição fundiária. Notamos a intensificação de projetos de lei que ameaçam a democratização da terra.	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1]
As mudanças climáticas são um dos maiores desafios da humanidade. O aumento da concentração de gases de efeito estufa (GEE), intensificado pela ação humana desde a Revolução Industrial e com enorme crescimento ao longo do século XX, impacta todo o mundo. Estes impactos são evidenciados pela maior frequência de eventos climáticos como secas, enchentes e ventos fortes, as alterações nos ciclos hidrológicos e as consequentes mudanças nos padrões de produtividade agrícola.	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0]

Tabela 1. *Exemplos de registros do conjunto de dados treinamento*.

O conjunto de dados foi aleatoriamente dividido em treinamento (80 %) e teste (20%). Do conjunto de treinamento foi separado 10% para validação.

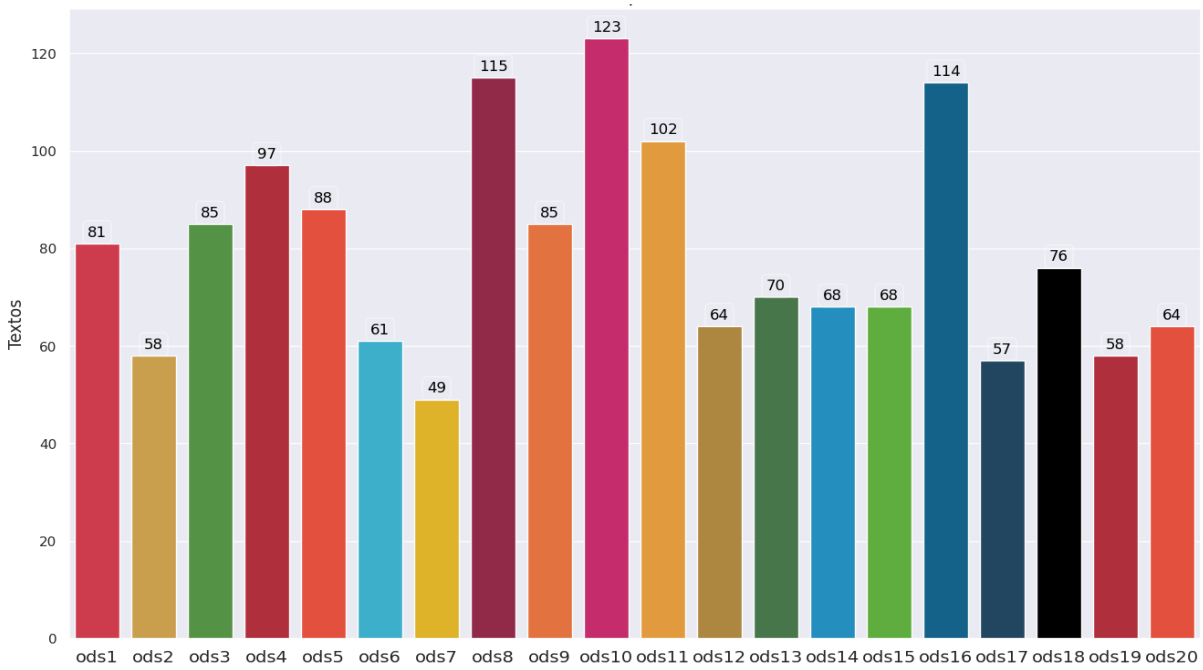

Figura 6*: Distribuição da quantidade de textos por ODS. Autor (2024)*

.

A figura 6 apresenta a distribuição dos textos coletados em relação aos 20 ODS. Por se tratar de uma classificação multirrótulo, o mesmo texto pode estar relacionado a mais de um ODS. Para equilibrar o treinamento e generalização do modelo foi utilizado pesos ponderados na função de perda.

4.1.2 AVALIAÇÃO DOS MODELOS PRÉ-TREINADOS

A linguagem de programação Python (== 3.10) foi usada para lidar com a estruturação de dados e avaliação dos modelos. Não foi realizado pré-processamento convencional nos dados brutos do corpus, preservando características estatísticas e linguísticas para facilitar a extração de recursos. Isso inclui a manutenção de maiúsculas e minúsculas, a retenção de pontuação e caracteres, bem como a preservação de stopwords, permitindo que o algoritmo identifique sinais importantes para a previsão.

Os experimentos foram realizados no programa Google Colab Pro + usando uma GPU V100, rodando em Ubuntu 22.04.3 LTS com aproximadamente 51GB de RAM disponível e 16GB de memória de vídeo.

A biblioteca de transformadores (== 4.41.2) desenvolvida por Hugging Face (Hugging Face, 2016), foi utilizada para ajustar os modelos pré-treinados Albertina PTBR100M (*PORTULAN/albertina-100m-portuguese-ptbr-encoder*), BERTimbau Base (*neuralmind/bert-base-portuguese-cased*) e DeBERTinha (*sagui-nlp/debertinha-ptbr-xsmall*). Esses modelos foram comparados com base nas métricas de desempenho selecionadas, e o modelo que apresentou o melhor desempenho foi então selecionado e ajustado para o desenvolvimento do ODSBahia-PTBR.

Através da biblioteca de otimização Bayesiana Optuna (== 3.6.1) (Akiba *et al*. 2019), foram realizadas múltiplas iterações para determinar os valores ideais dos hiperparâmetros. Após vários experimentos, foram considerados os valores otimizados para os hiperparâmetros: o número de iterações (*epochs*, variou de 2^3 a 2^5), o tamanho do lote (variou de 2^2 a 2^4), a taxa de aprendizado que variou de aproximadamente 10^{-5} a 10^{-4}. Como função de perda, foi utilizada a Entropia Binária Cruzada com perda de logit (*BCEWithLogitsLoss*) com ajuste de pesos (*pos_weight*) e o algoritmo de otimização AdamW (Loshchilov; Hutter, 2019).

4.2 Inferência sobre as ações universitárias

Conforme apresentado na Figura 7, inicialmente foi conduzido a etapa de seleção das instituições, onde foram selecionadas três universidades federais do Estado da Bahia devido à sua importante atuação regional e ao potencial impacto de suas atividades de pesquisa e extensão no contexto dos ODS. Na Bahia, estão presentes seis universidades públicas federais, a Universidade Federal do Bahia (UFBA), Universidade Federal do Oeste da Bahia (UFOB), Universidade Federal do Recôncavo da Bahia (UFRB), Universidade Federal do Sul da Bahia (UFSB), Universidade Federal da Lusofonia Afro-Brasileira (Unilab), Universidade Federal do Vale do São Francisco (Univasf) (BRASIL, 2024).

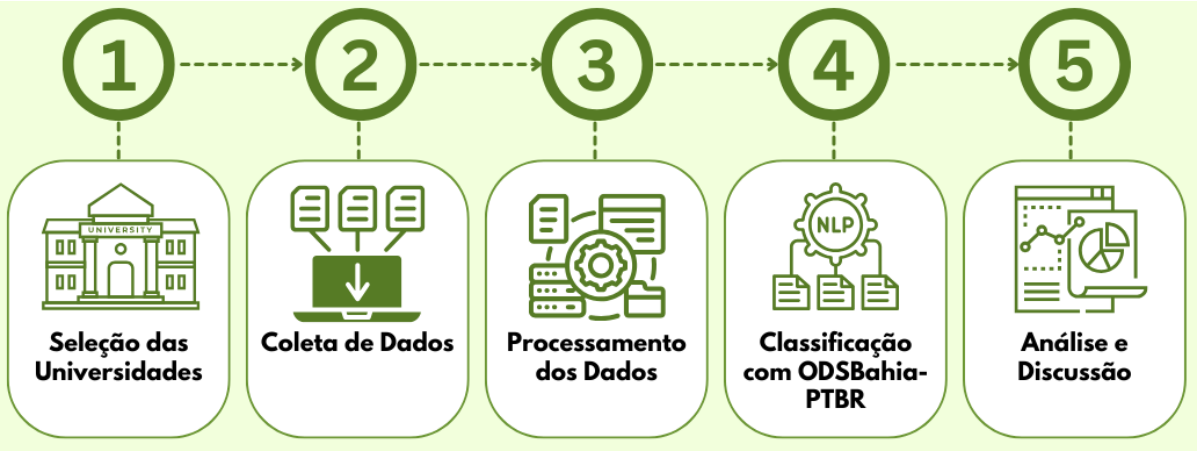

***Figura** 7: Metodologia Analítica para análise das ações universitárias. Autor (2024)*

Contudo, o critério de seleção considerou, além da relevância regional das instituições, o uso do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) como sistema de gestão acadêmica e a disponibilização pública de suas atividades e projetos nos anos de 2019 a 2023 por meio desse sistema. Das seis instituições, apenas três possuíam registros disponíveis publicamente no SIGAA que abrangiam o período proposto.

Cada uma das instituições selecionadas, a UFSB, UFRB e UFOB, está localizada em áreas distintas do estado da Bahia (Sul, Recôncavo e Oeste), com características socioeconômicas e ambientais variadas que influenciam suas estratégias e prioridades nas ações para o desenvolvimento sustentável. Além disso, elas utilizam como sistema de gestão acadêmica o SIGAA, possibilitando uma análise consistente e comparativa dos dados. Inicialmente, foi realizada uma pesquisa documental, que incluiu a coleta e análise de registros de ações desenvolvidas. Adicionalmente, foram utilizados documentos oficiais sobre os ODS,

dados do Instituto Brasileiro de Geografia e Estatística (IBGE) e informações fornecidas publicamente pelas próprias instituições de ensino.

Em seguida, foi conduzido um levantamento da literatura existente sobre os temas norteadores desta pesquisa, incluindo o uso e aplicação de processamento de linguagem natural no contexto dos ODS, bem como a educação para a sustentabilidade e a Agenda 2030 nas instituições de ensino superior. Este levantamento teve como objetivo contextualizar e fundamentar teoricamente o estudo, a análise e discussão dos resultados.

A amostra de dados analisada inclui os dados de projetos de pesquisa e extensão que foram extraídos da página pública do SIGAA da UFSB, UFRB e UFOB por meio da técnica de *'Webscraping'*, garantindo precisão e abrangência. Esse processo envolveu a extração de conteúdo e rótulos dos textos por meio da análise da codificação HTML do sistema.

A coleta da amostra de dados foi realizada entre os dias 02 e 08 de janeiro de 2024, em duas etapas. A primeira consistiu na identificação e extração dos registros, utilizando filtros de busca no SIGAA. Foram selecionados todos os dados disponíveis, incluindo: título do projeto, resumo, localidade, tipo do projeto, área de enfoque e palavras-chave. Na segunda etapa, os dados foram exportados para um formato que permite análise computacional (planilha CSV). Para automatizar a coleta, foi desenvolvido um algoritmo em Python (== 3.6).

A análise dos dados foi conduzida a partir do ODSBahia-PTBR que permitiu a identificação e categorização dos ODS relacionados às ações acadêmicas.

Os seguintes passos foram seguidos:

1. Pré-processamento dos dados, onde foram realizadas limpeza (remoção de registros duplicados ou explicitamente cadastrados como testes) e preparação dos textos dos resumos. Nessa etapa, os projetos foram organizados, municípios identificados e removidos projetos que não possuíam textos em seus resumos.
2. Classificação automática utilizando o classificador ODSBAHIA-PTBR para identificar os ODS relacionados a cada projeto.
3. Análise descritiva, por meio de cálculos de frequências e distribuições dos ODS associados aos projetos das três universidades.

A pesquisa respeitou todas as normas éticas aplicáveis, considerando se tratar de informações públicas disponibilizadas pelas instituições. Nenhuma informação pessoal identificável foi utilizada na análise, e todos os dados foram tratados de forma anônima e confidencial, garantindo a privacidade e a integridade das informações. Para as análises, além dos 17 ODS da ONU, incluiremos também o ODS 18 (Igualdade Étnico-Racial) como parte do

escopo, pois ele já está incorporado à Agenda 2030 nacional e é relevante para abordar questões locais e específicas que não estão totalmente contempladas pelos ODS originais da ONU.

4.2.1 ÁREA DE ESTUDO - CONTEXTO DAS INSTITUIÇÕES ANALISADAS

A seguir, são descritas algumas das principais especificidades de cada universidade selecionada e os municípios em que estão inseridas, a fim de permitir uma compreensão do cenário dessas instituições em seus territórios de atuação.

4.2.1.1 UNIVERSIDADE FEDERAL DO SUL DA BAHIA

A UFSB foi criada em 2013 pela sanção presidencial da Lei nº 12.818/2013 e iniciou suas atividades em setembro de 2014 com o ingresso das primeiras turmas. A UFSB abrange os municípios situados em três dos 27 Territórios de Identidade da Bahia, especificamente Litoral Sul, Costa do Descobrimento e Extremo Sul, ocupando 40.384 km² e englobando 48 municípios, numa região caracterizada por uma rica diversidade cultural, social e ambiental.

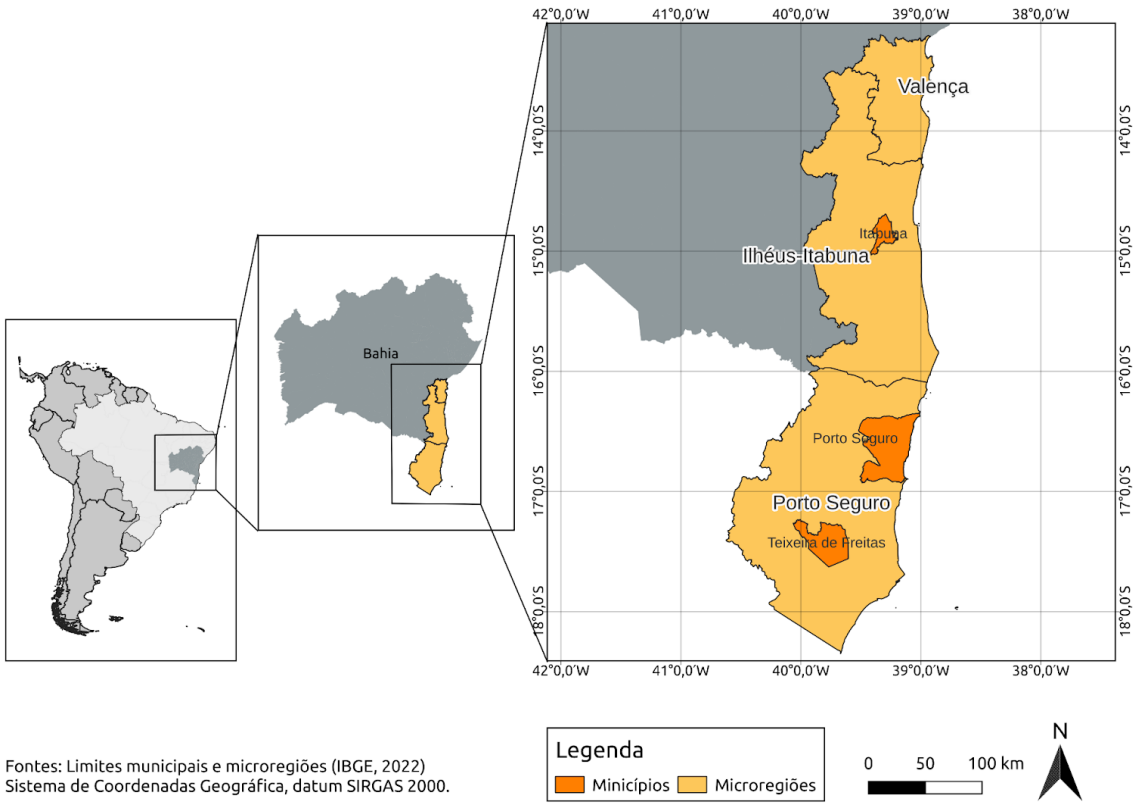

***Figura 8**: Mapa de localização dos três campi da UFSB no território sul da Bahia. O mapa destaca a presença dos campi da UFSB nos municípios de Porto Seguro, Teixeira de Freitas e Itabuna, onde também está a Reitoria. Autor (2024)*

Os três campi da UFSB estão localizados nos municípios de Itabuna (Campus Jorge Amado - CJA), onde também está localizada a Reitoria, Porto Seguro (Campus Sosígenes Costa - CSC) e Teixeira de Freitas (Campus Paulo Freire - CPF), os quais possuem uma população estimada de 186,7 mil, 168,3 mil e 145,2 mil habitantes, respectivamente (Figura 8).

Além disso, a instituição possui uma abrangência ampliada por meio da Rede Anísio Teixeira de Colégios Universitários (Rede CUNI), que é um projeto da UFSB em parceria com o Governo do Estado da Bahia, que busca promover a inclusão social por meio da educação superior através de unidades acadêmicas situadas em instalações da rede estadual de educação. Os CUNIs estão presentes em outros 7 municípios do território: Ilhéus (178.649), Coaraci (17.333), Ibicaraí (21.665), Santa Cruz Cabrália (29.185), Eunápolis (113.710), Itamaraju (59.605), Nova Viçosa (39.509). Esses municípios, juntos, têm uma população de aproximadamente 459,6 mil habitantes (IBGE, 2022. UFSB, 2022).

A UFSB está inserida em uma região marcada por grande diversidade ambiental, social e cultural. Seu território é caracterizado pela presença predominante do bioma Mata Atlântica, com áreas de relevante interesse ecológico e zonas costeiras sensíveis. Essa região abriga diversas comunidades tradicionais, incluindo povos indígenas, quilombolas, e pescadores artesanais, o que contribui para a complexidade sociocultural e os desafios da sustentabilidade local (UFSB, 2019).

O clima predominante nos territórios de atuação da UFSB varia entre úmido e subúmido, com temperaturas médias elevadas e chuvas concentradas nos primeiros meses do ano. As atividades econômicas da região incluem o turismo (com destaque para Porto Seguro e Santa Cruz Cabrália), a agricultura familiar, a silvicultura, a produção de cacau e o setor de serviços. A presença de áreas de preservação ambiental, como parques e reservas, também evidencia o potencial e a necessidade de integração entre desenvolvimento regional e conservação ambiental (SANTOS, 2022; UFSB, 2019).

Este contexto territorial confere à UFSB um papel estratégico na formulação e implementação de ações acadêmicas que contribuam para o desenvolvimento sustentável e a redução das desigualdades regionais, alinhando-se aos objetivos propostos pela Agenda 2030.

4.2.1.2 UNIVERSIDADE FEDERAL DO RECÔNCAVO BAIANO

A Universidade Federal do Recôncavo da Bahia (UFRB) foi oficializada pela Lei nº. 11.151 em 2005 e inaugurada em 2006 pelo presidente Luiz Inácio Lula da Silva. A UFRB, concebida como modelo multicampi, conforme destacado na Figura 9, está localizada nas

cidades de Amargosa (36.521), Cachoeira (29.250), Feira de Santana (616.272), Santo Amaro (56.012), Santo Antônio de Jesus (103.055) e São Félix (11.026) e tem sua Reitoria localizada em Cruz das Almas (60.348), a 146 quilômetros da capital do estado. Esses municípios, juntos, têm uma população de aproximadamente 912,4 mil habitantes (IBGE, 2022. UFRB, 2019).

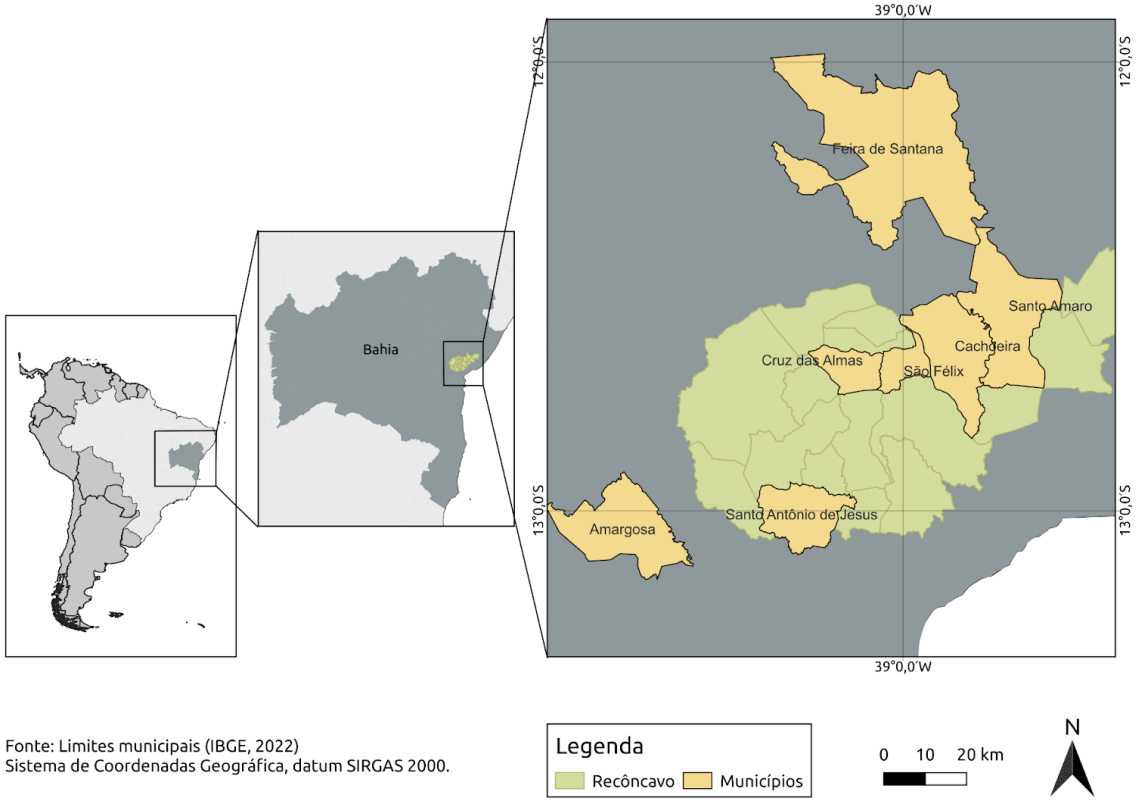

***Figura 9**: Mapa de localização da UFRB no Recôncavo baiano. Destaca-se a presença da UFRB em sete municípios: Amargosa, Cachoeira, Feira de Santana, Santo Amaro, Santo Antônio de Jesus, São Félix e tem sua Reitoria localizada em Cruz das Almas. Autor (2024)*

O Território do Recôncavo é uma área composta por 20 municípios, ocupa uma área de 5.250,51 km², e está localizado em torno da Baía de Todos-os-Santos, incluindo tanto áreas litorâneas quanto regiões interioranas próximas. As faixas climáticas são bem distribuídas entre os tipos subúmido e seco no oeste, úmido a subúmido na porção central e úmido, no leste. Na economia, destaca-se o setor de comércio e serviços, seguido pela indústria e agropecuária. O território mantém variadas tradições culturais indígenas e de matrizes africanas que contribuem para a formação da identidade da cultura baiana. Seu patrimônio arquitetônico, festas populares e religiosidade revelam seu potencial turístico (UFRB, 2022).

Essa diversidade não apenas enriquece o ambiente acadêmico da UFRB, mas também propicia um campo fértil para atividades acadêmicas que buscam promover o desenvolvimento sustentável.

4.2.1.3 UNIVERSIDADE FEDERAL DO OESTE DA BAHIA

A Universidade Federal do Oeste da Bahia (UFOB) foi criada pela Lei nº 12.825, de 5 de junho de 2013, com início das atividades acadêmicas em 2014, a partir de um desmembramento do Instituto de Ciências Ambientais e Desenvolvimento Sustentável (ICADS) da Universidade Federal da Bahia.

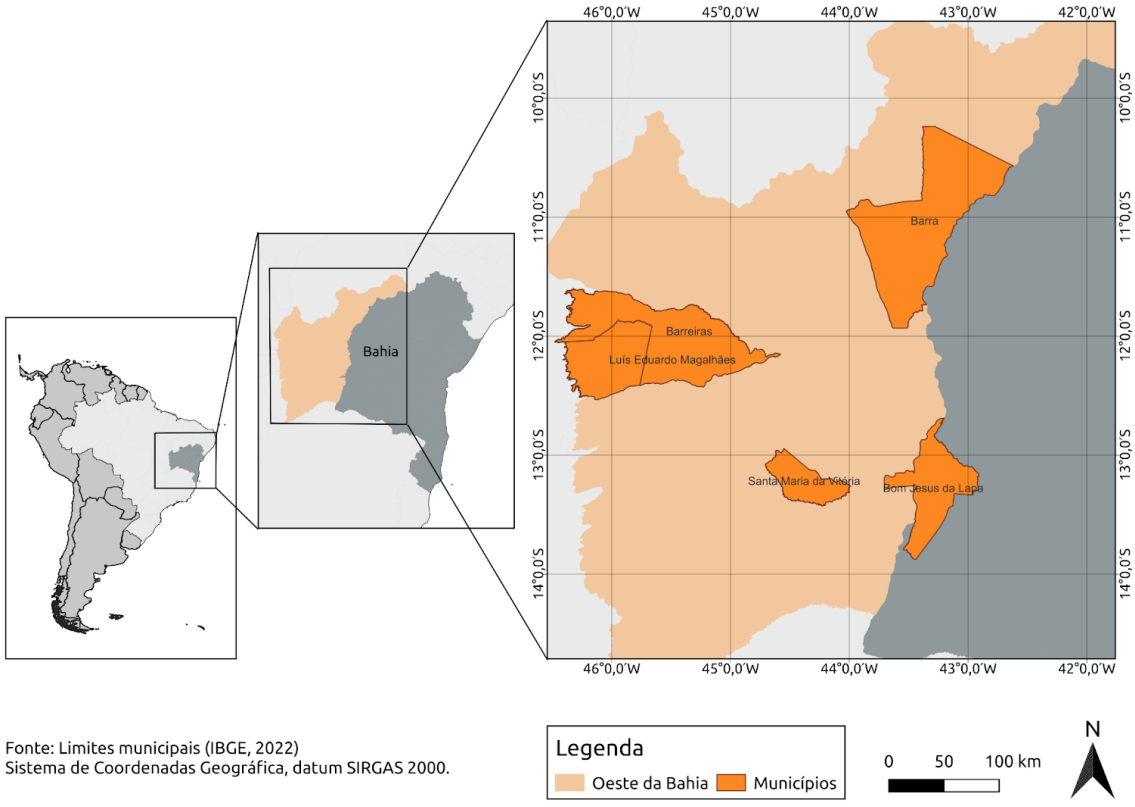

***Figura 10**: Mapa de localização da UFOB no Oeste da Bahia, destacando sua presença em cinco municípios. Autor (2024)*

A Figura 10, traz em destaque as cidades de localização da UFOB. Com sua sede em Barreiras (159.734), a UFOB está presente também em outros quatro municípios: Barra (51.092), Bom Jesus da Lapa (65.550), Luís Eduardo Magalhães (107.909) e Santa Maria da Vitória (38.604). Esses municípios, juntos, têm uma população de aproximadamente 422,8 mil habitantes. Ao todo, o território de atuação da UFOB abrange uma área contínua de 223 mil

km², considerando um raio de 150 quilômetros a partir das sedes dos municípios que abrigam suas unidades. (IBGE, 2022. UFOB, 2023).

O Oeste da Bahia se destaca por suas características naturais marcantes, como o clima Tropical Continental, com estações bem definidas – verão chuvoso e inverno seco – e áreas com clima Tropical Semiárido. O relevo é composto pela Depressão Sertaneja e pelo Chapadão Ocidental do São Francisco, enquanto a rede hidrográfica pertence à Bacia do Rio São Francisco. Essas condições naturais, aliadas a práticas agrícolas, transformaram a região em um importante polo produtor de grãos, como soja e milho, além de culturas como o algodão (AIBA, 2023).

5 RESULTADOS E DISCUSSÕES

5.1 Avaliação do modelo ODSBAHIA-PTBR

Os resultados apresentados nesta seção referem-se ao desempenho dos modelos pré-treinados em relação ao conjunto de teste. As métricas de precisão, *recall*, e a micro *F1-Score* foram calculados utilizando *scikit-learn* (== 1.3.2) com o limiar de 0,5. Com base nos resultados obtidos (Figura 11), houve pequenas diferenças no desempenho dos modelos nesta tarefa de classificação multirrótulo para os 20 ODS, mas que podem impactar nas inferências.

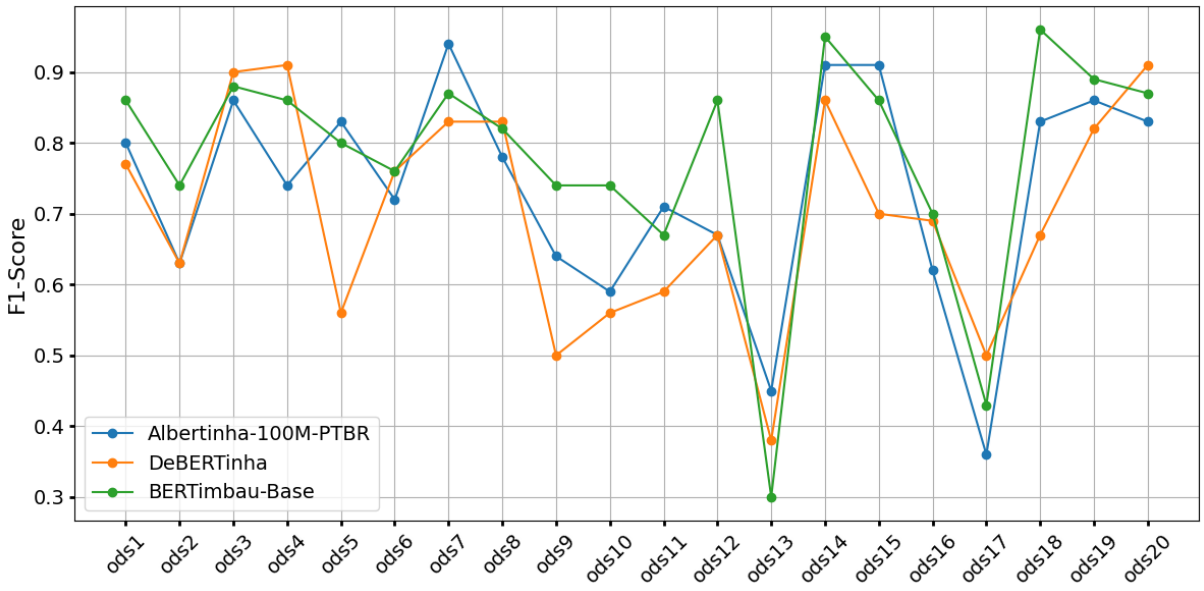

***Figura 11**: F1-Score por ODS para cada modelo pré-treinado. Autor (2024)*

Os modelos apresentaram desempenhos muito semelhantes em termos de *F1-Score*, o que sugere que todos eles são eficazes em manter um equilíbrio nas previsões. BERTimbau

Base e Albertina 100M PTBR tiveram um desempenho um pouco melhor em comparação com DeBERTinha. Contudo, DeBERTinha apesar de ter um *recall* mais baixo (Tabela 2), ainda mantém um *F1-Score* competitivo, indicando que também seria uma escolha viável. O BERTimbau Base foi selecionado e utilizado como base para ajustar o modelo ODSBahia-PTBR.

Modelo	**Precision**	**Recall**	***F1-Score***
BERTimbau Base	0.89	0.70	0.79
Albertina 100M PTBR	0.80	0.71	0.75
DeBERTinha	0.84	0.63	0.72

Tabela 2 - *Resultados dos ajustes nos modelos*.

O BERTimbau Base mostrou uma precisão de 0,89, o que significa que a maioria das previsões feitas pelo modelo está correta. Esse valor é importante para reduzir erros na classificação de textos que não estão alinhados com os ODS. O *recall* de 0,70 sugere que o modelo consegue identificar corretamente 70% dos ODS relevantes, embora não tenha conseguido identificar os ODS em 30% dos casos. O F1-Score de 0,79 indica um bom equilíbrio entre precisão e *recall*, sugerindo que o modelo consegue identificar textos que realmente pertencem ao ODS em questão quanto em manter a exatidão (garantir que os textos identificados como pertencentes ao ODS realmente o sejam).

Isolando as predições do modelo multirrótulo para cada ODS individualmente, como em um problema de classificação binária, os seguintes resultados do BERTimbau Base foram obtidos:

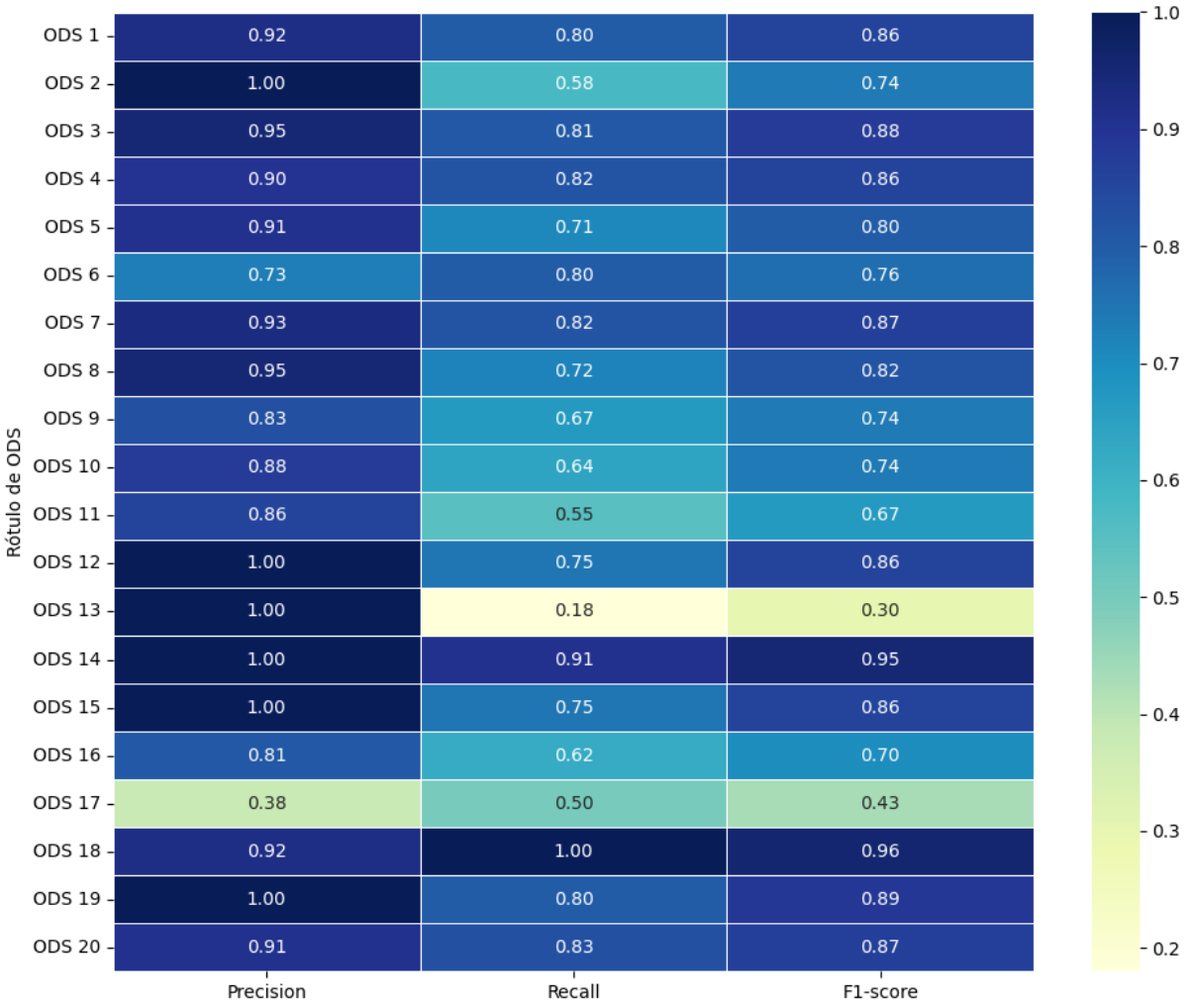

***Figura 12**: Métricas binárias para cada ODS. Autor (2024)*

Observa-se um bom desempenho em alguns rótulos, como ODS 3, ODS 5, ODS 14, ODS 18 e ODS 20, onde as métricas de precisão e *recall* são altas, resultando em uma *F1-Score* elevada. Por outro lado, há espaço para melhorias em alguns ODS, como os 6, 11, 13 e 17, onde o *recall* é mais baixo. Além disso, alguns rótulos apresentaram um desequilíbrio entre precisão e *recall,* como o ODS 2 e o ODS 16. Os ODS 13 e 17 apresentaram os menores resultados, possivelmente devido à complexidade dos temas e à falta de exemplos de treinamento suficientes que cobrem adequadamente as nuances desses objetivos.

Recentemente, diversos estudos têm explorado técnicas de PLN em várias aplicações relacionadas aos ODS. Fux *et al*. (2022) desenvolveram uma ferramenta para classificação de processos judiciais e sua correlação com os ODS. Para isso, coletaram documentos em PDF de petições iniciais e de acórdãos de processos do STF, e extraíram textos utilizando a técnica de reconhecimento óptico de caracteres (OCR). O estudo, que utilizou técnicas de pré-

processamento de texto e criação de *embeddings*, empregou redes neurais recorrentes juntamente com outras abordagens clássicas, como vetores binários (*one-hot encoding*) e representações de frequências de palavras (*Bag-of-Words* - *BoW* e *TF-IDF*).

Chen *et al*. (2023) conduziram um estudo com PLN voltado à governança ambiental, social e corporativa (*Environmental, Social and Governance* - ESG), a fim de identificar empresas alinhadas com os ODS, com base no texto de suas divulgações de sustentabilidade. Analisando relatórios de responsabilidade social corporativa de empresas do índice Russell 1000 (índice do mercado de ações dos EUA) entre 2010 e 2019, aplicaram classificadores logísticos, máquinas de vetores de suporte e uma rede neural totalmente conectada para prever o alinhamento com os ODS. Especificamente, usaram embeddings de palavras como recurso de entrada baseados em dicionários, baseados nos modelos *Word2Vec* e *Doc2Vec*, para classificar o alinhamento das empresas com os ODS ao longo do tempo.

Kharlashkin *et al*. (2023) apresentam um estudo sobre o alinhamento de cursos universitários em relação aos ODS. Para isso, eles utilizaram o modelo de linguagem PaLM 2 para gerar dados de treinamento a partir de descrições desses cursos e usaram os dados gerados para treinar alguns modelos de linguagem menores visando a previsão dos ODS. Entre os modelos pré-treinados utilizados estão o BERT, mBERT, RoBERTa, XLM-RoBERTa e BART, que foram ajustados para tarefas de classificação multiclasse. Os autores buscaram contribuir para uma melhor compreensão dos ODS no ambiente universitário, permitindo a identificação de como a Agenda 2030 está alinhada a cada curso.

Alguns autores se concentraram na tarefa de classificação multirrótulo utilizando documentos em inglês relacionados aos ODS. Guisiano, Chiky & Mello (2021) conduziram uma pesquisa na qual desenvolveram uma ferramenta de classificação multirrótulo para textos, empregando o modelo BERT. Para isso, eles coletaram textos já rotulados relacionados a um ou mais ODS do site IISD SDG Knowledge Hub, assegurando que cada texto contivesse um mínimo de 512 palavras. A coleta foi realizada por meio da técnica de *Webscraping*. Os autores utilizaram a biblioteca *Fast-Bert11*, que é construída sobre a base do *PyTorch* e suporta os modelos BERT e XLNet, para realizar a classificação de texto tanto em cenários multiclasse quanto em multirrótulo. Essa abordagem permitiu uma análise com precisão de 98% sob uma quantidade limitada de dados de treinamento.

Callaghan *et al*. (2021) utilizaram DistilBERT com validação cruzada aninhada para categorizar e extrair informações específicas de 102.160 estudos sobre impactos climáticos, focando no ODS 13. Usando classificadores multirrótulo para prever categorias de impacto e

drivers climáticos nos documentos incluídos, os autores utilizaram os resultados para criar um banco de dados e mapear tendências em mudanças climáticas antropogênicas no período de 1951 a 2018.

O estudo conduzido por Matsui *et al*. (2022) também no campo da classificação multirrótulo dos ODS utilizando o modelo BERT, envolveu o ajuste de um modelo pré-treinado pela Wikipedia japonesa, a variante base do Tohoku-BERT (*cl-tohoku/bert-base-japanese-whole-word-masking*) com o objetivo de identificar relação com ODS em sentenças de entrada, e vetorização da semântica das sentenças.

Analisando os resultados juntamente com os estudos mencionados anteriormente, as métricas obtidas pelo modelo ajustado ODSBahia-PTBR indicam um desempenho similar, conforme apresentado na Tabela 3.

Propostas	**Precision**	**Recall**	***F1-Score***
Matsui *et al.* (2022)	0.93	0.92	0.93
Kharlashkin *et al.* (2024)	0.769	0.803	0.786
Este modelo (ODSBahia-PTBR)	**0.89**	**0.70**	**0.79**
Chen *et al.* (2021)	0,819	0,744	0,745
Callaghan *et al.* (2021) (ODS 13)	-	-	0.71
RAFA 2030 (Fux *et al.*, 2022) (ODS 3,8,10,16)	-	0.881	0.636

Tabela 3 - *Métricas de classificadores de texto em relação aos ODS*.

A fim de avaliar o desempenho em um conjunto de dados maior, a base de dados pública OSDG Community Dataset (OSDG-CD) foi traduzida e submetida ao classificador ODSBahia-PTBR. A tradução do inglês para o português foi executada por meio de uma Interface de Programação de Aplicativos (API, do inglês, *Application Programming Interface*), a qual viabilizou o envio de requisições de tradução à plataforma online Google Translate.

OSDG-CD é um conjunto de dados público com milhares de trechos de texto de parágrafos derivados de mais de três mil documentos disponíveis publicamente. Todos esses textos são rotulados com o ODS relacionado e foram validados por mais de 1.400 voluntários de mais de 140 países. Cada texto é composto de 3 a 6 frases e tem em média cerca de 90 palavras. O arquivo de conjunto de dados CSV utilizado (versão 2024.01 disponibilizado em 01 de janeiro de 2024) conta com 42.630 trechos de texto e um total de 306.595 rótulos de ODS, exceto o ODS 17 (OSDG, 2024). Cada amostra fornece uma pontuação de consistência calculada a partir dos resultados das anotações dos diferentes voluntários. Somente as amostras cujo "*labels_positive*" > "*labels_negative*" foram utilizadas, retendo 35.218 amostras.

O ODSBahia-PTBR foi capaz de, para além de acertar o rótulo corretamente, identificar demais correlações presentes nos textos da OSDG-CD. Analisando individualmente cada

rótulo, a partir de uma classificação binária, o ODSBahia-PTBR obteve uma acurácia de 0.9523 na classificação dos rótulos de ODS 1 a 16 (Figura 13).

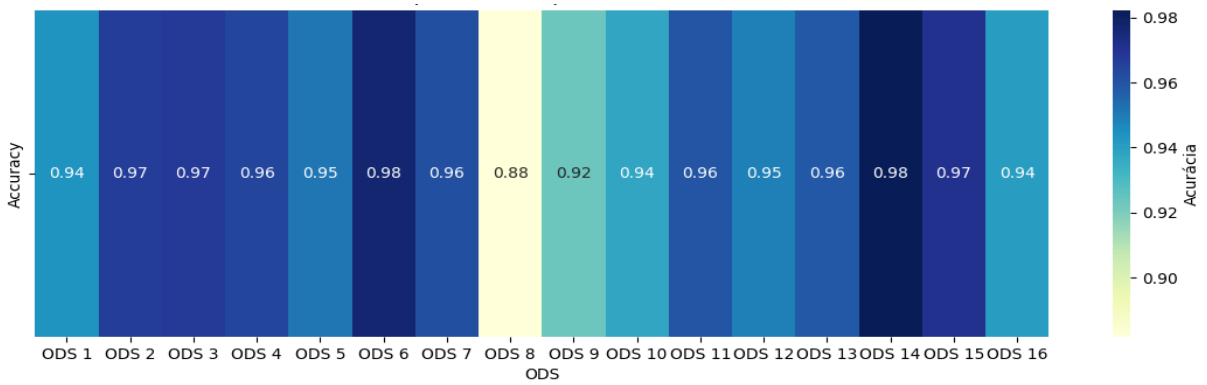

***Figura 13**: Acurácia por ODS sobre a base OSDG-CD. Autor (2024).*

Os estudos recentes, discutidos nesta seção, demonstram a importância e aplicabilidade do uso de técnicas de PLN para apoiar os ODS. Essa interseção entre inteligência artificial e ODS abre uma gama de oportunidades para monitorar e abordar os desafios do desenvolvimento sustentável de forma mais autônoma.

Estudos como os de Fux *et al*. (2022), Chen *et al*. (2023) e Kharlashkin *et al*. (2023) exemplificam a aplicação prática do PLN em diferentes contextos, desde a classificação de processos judiciais até a avaliação de relatórios de sustentabilidade corporativa e a análise de cursos universitários. Esses trabalhos também ressaltam a importância de considerar as nuances linguísticas e culturais ao adaptar modelos para diferentes idiomas.

A construção de um corpus em português para o modelo ajustado ODSBahia-PTBR, preenche uma lacuna na literatura, possibilitando uma análise mais contextualizada dos ODS. A abordagem apresentada aqui, pode ser ampliada e adaptada para diferentes propósitos, como, por exemplo, a avaliação de políticas públicas, o monitoramento de iniciativas empresariais, e a promoção de práticas sustentáveis em comunidades locais.

Além disso, a proposta deste trabalho de abranger a classificação de textos também em relação aos três novos ODS voltados para Igualdade Racial (ODS 18), Arte, Cultura e Comunicação (ODS 19) e Direitos dos Povos Originários e Comunidades Tradicionais (ODS 20), permite que o modelo ODSBahia-PTBR amplie a capacidade de identificar iniciativas que impactam esses grupos.

Portanto, este modelo, especificamente ajustado para o contexto brasileiro, foi desenvolvido para compreender e processar nuances linguísticas e culturais diversas, ampliando

sua aplicabilidade não apenas aos 17 ODS da ONU, mas também a esses três adicionais propostos.

5.2 Como as universidades fazem o que falam sobre a Agenda 2030?

A análise das iniciativas acadêmicas das universidades UFSB, UFRB e UFOB revelou informações importantes sobre a relação entre ensino, pesquisa, extensão e os ODS. A categorização dos projetos utilizando PLN permitiu correlacionar as atividades locais com os ODS de maneira objetiva e autônoma, e identificar como essas universidades operacionalizam seu compromisso com a Agenda 2030. A seguir, apresentamos os principais achados, organizados por frequência de ocorrência dos ODS, padrões de coocorrência, distribuição geográfica das iniciativas e lacunas identificadas.

Inicialmente, foram coletados 7.013 registros dos projetos de pesquisa e extensão ao todo, sendo eles categorizados pelas instituições como projetos de pesquisa interno ou externo, e ações de extensão como sendo cursos, eventos, projetos, produtos, programas e publicações ou produções didáticas. Após a eliminação dos registros que não continham texto no campo resumo, a análise foi realizada com um total de 6.983 registros, distribuídos temporalmente conforme **tabela 4**.

Instituição	***P***	***E***	***P***	***E***	***P***	***E***	***P***	***E***	***P***	***E***	***Total***
UFOB	*144*	*29*	*74*	*154*	*78*	*132*	*81*	*149*	*139*	*211*	*1191*
UFRB	*63*	*1*	*77*	*11*	*152*	*1050*	*211*	*1140*	*191*	*1438*	*4334*
UFSB	*128*	*145*	*137*	*107*	*125*	*151*	*134*	*184*	*141*	*206*	*1458*
	2019		***2020***		***2021***		***2022***		***2023***		***6983***

P = pesquisa; E = extensão;

Tabela 4: *Distribuição dos projetos de pesquisa (P) e ações de extensão (E) realizados entre 2019 e 2023 coletados nas instituições analisadas.*

5.2.1 INFERÊNCIA

O processo de inferência foi realizado utilizando o modelo ODSBAHIA-PTBR (Santos; Moraes, 2024). Durante esta etapa, os resumos dos projetos de pesquisa e extensão foram utilizados como entradas no modelo. O texto é inicialmente tokenizado e segmentado em partes menores, respeitando o limite de 512 *tokens* por segmento que é processado individualmente pelo modelo. O modelo captura padrões linguísticos e semânticos importantes através de suas camadas internas que são responsáveis por gerar representações (*embeddings*) contextuais para

os *tokens*. Essas representações geradas são então mapeadas para um espaço de saída correspondente aos diferentes ODS. As saídas do modelo são convertidas em probabilidades utilizando a função *sigmoide*.

Depois de obter as probabilidades dos segmentos do texto, calculamos a média dessas probabilidades para chegar a uma previsão final, incluindo apenas aquelas que excedem o limiar de 0,2, permitindo que o modelo esteja mais sensível. Isso significa que qualquer ODS inferido com uma probabilidade igual ou superior a esse valor foi considerado relevante.

Dessa forma, com base nas probabilidades médias e nos padrões linguísticos e semânticos identificados, o modelo inferiu quais ODS estavam relacionados a cada projeto, permitindo uma classificação automatizada e consistente dos dados. Por ser um classificador multirrótulo, o ODSBAHIA-PTBR pôde associar um mesmo texto a um ou mais ODS. Como resultado, o histograma ilustrado na Figura 14 mostra a frequência de ocorrências de cada ODS na amostra.

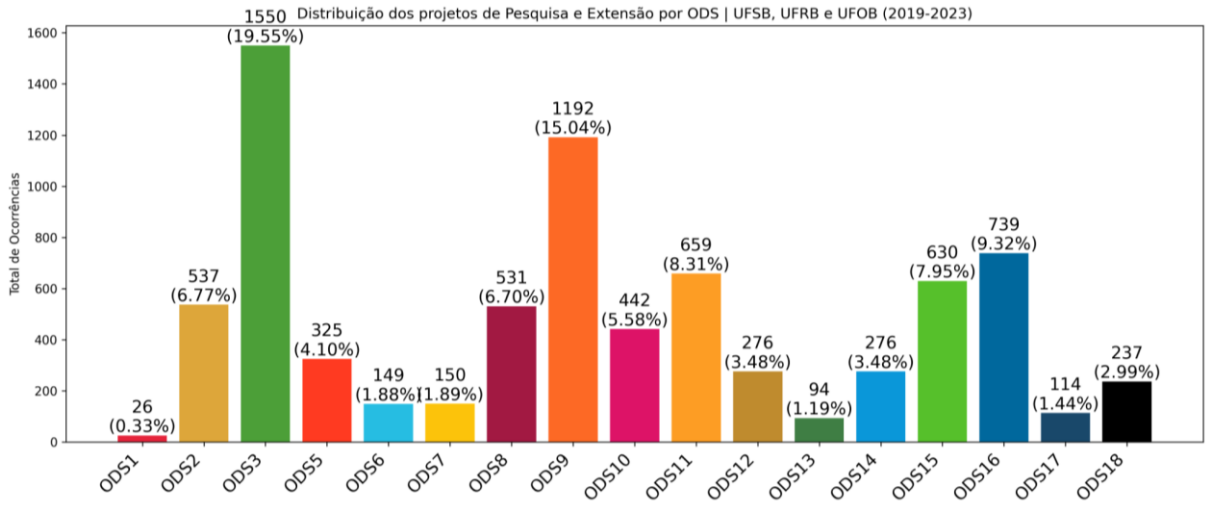

***Figura 14**: Frequência de ocorrências dos ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024)*

Removemos a indicação do ODS 4 ("Educação de Qualidade") no gráfico das figuras 14, 15, 19, 22 e 23 para deixar a visualização mais clara e equilibrada. A promoção da educação de qualidade é um compromisso fundamental e intrínseco das universidades, de modo que, este objetivo apresentou o maior número (4.226) de alinhamentos nas ações das universidades analisadas e sua inclusão no gráfico provocou um desbalanceamento na comparação com os outros ODS.

Observa-se que a distribuição de projetos não é uniforme entre os ODS. Alguns objetivos, como o ODS 3 (Saúde e Bem-Estar), ODS 9 (Indústria, Inovação e Infraestrutura),

ODS 16 (Paz, Justiça e Instituições Eficazes) e ODS 11 (Cidades e Comunidades Sustentáveis), têm uma quantidade significativa de coocorrências que contribuem para estes objetivos, sendo majoritariamente atendidos por projetos de extensão (Figura 15). Outros, como o ODS 1 (Erradicação da Pobreza), ODS 13 (Ação Contra a Mudança Global do Clima), ODS 17 (Parcerias e Meios de Implementação), ODS 6 (Água Potável e Saneamento) e ODS 7 (Energia Acessível e Limpa), o modelo não foi capaz de identificar muitos relacionamentos.

Essa diferenciação entre pesquisa e extensão sugere que as universidades utilizam abordagens complementares para atender às demandas da sociedade, combinando pesquisa aplicada com ações diretas de impacto na comunidade local.

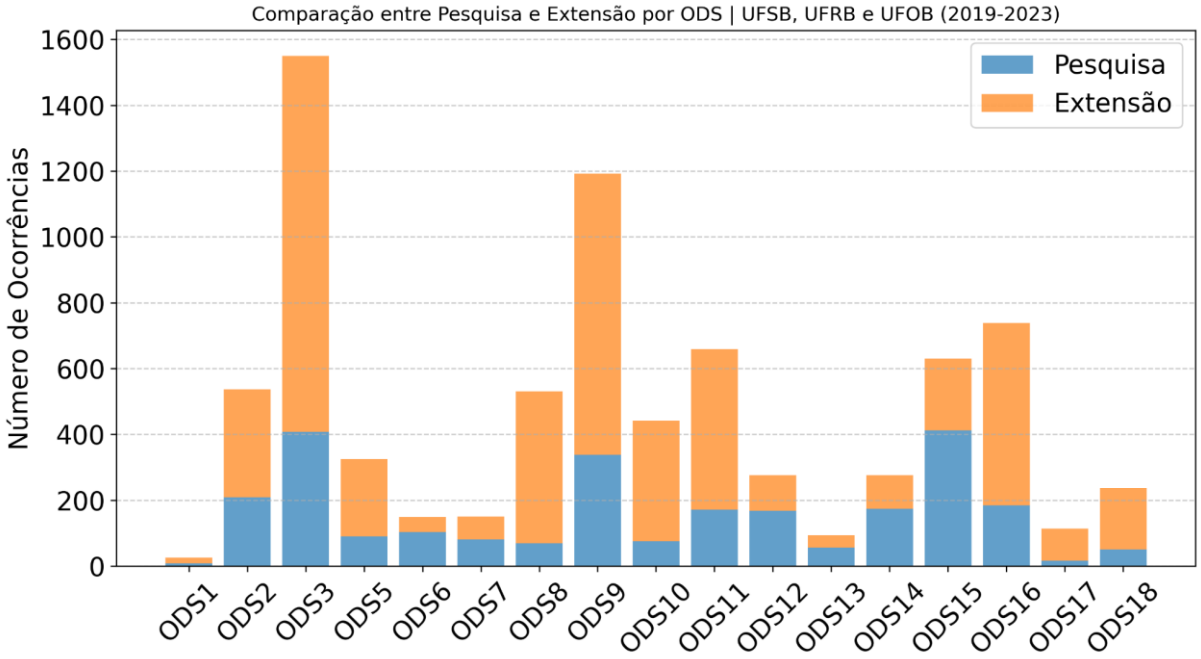

***Figura 15**: Comparação da frequência de ocorrências dos ODS entre a pesquisa e extensão. Autor (2024)*

A alta ocorrência do ODS 9 nos projetos analisados, indica que o ODSBAHIA-PTBR inferiu uma forte orientação para a pesquisa aplicada, desenvolvimento tecnológico e inovação. Esse ODS está focado em construir infraestruturas resilientes, promover a industrialização inclusiva e sustentável e fomentar a inovação (IPEA, 2019). A meta 9.5, por exemplo, tanto no contexto global quanto especificamente para o Brasil, enfatiza a importância de fortalecer a pesquisa científica e melhorar a capacidade tecnológica, especialmente visando o aumento da produtividade e inovação.

"Brasil - Fortalecer a pesquisa científica e melhorar as capacidades tecnológicas das empresas, incentivando, até 2030, a inovação, visando aumentar o emprego do conhecimento científico e tecnológico nos desafios socioeconômicos nacionais e nas tecnologias

socioambientalmente inclusivas; e aumentar a produtividade agregada da economia. [...]" (IPEA, 2019).

Embora a inovação seja um dos eixos do ODS 9, Galvão, Freitas e Menezes (2024) apontam que, no Brasil, o desenvolvimento sustentável baseado em inovação foi negligenciado nos últimos anos, principalmente pela falta de políticas públicas eficazes e pelo baixo investimento governamental. Fatores como recursos limitados, alto custo do crédito e baixos investimentos privados na indústria da inovação dificultam a implementação de uma cultura de inovação mais massiva no país. Em 2022, por exemplo, os investimentos públicos em pesquisa e desenvolvimento (P&D) representaram cerca de 0,62% do PIB, valor considerado insuficiente para impulsionar a inovação tecnológica no Brasil (MCTI, 2023).

As universidades são responsáveis por grande parte da pesquisa científica e do desenvolvimento tecnológico do país (UNESCO, 2021). O fato de o ODS 9 estar entre os mais presentes nas instituições analisadas aqui indica que o ensino superior tem sido um agente compensador da baixa prioridade dada à inovação em outros setores. As universidades frequentemente colaboram com empresas de diferentes setores para aplicar conhecimento técnico e científico, através de programas de cooperação técnica, estágios e parcerias (Vide ODS 16), isso possibilita que o conhecimento gerado nas universidades seja passado para outros atores, melhorando as capacidades tecnológicas e incentivando inovações práticas (estímulo à implementação de novas ideias e técnicas) nas empresas.

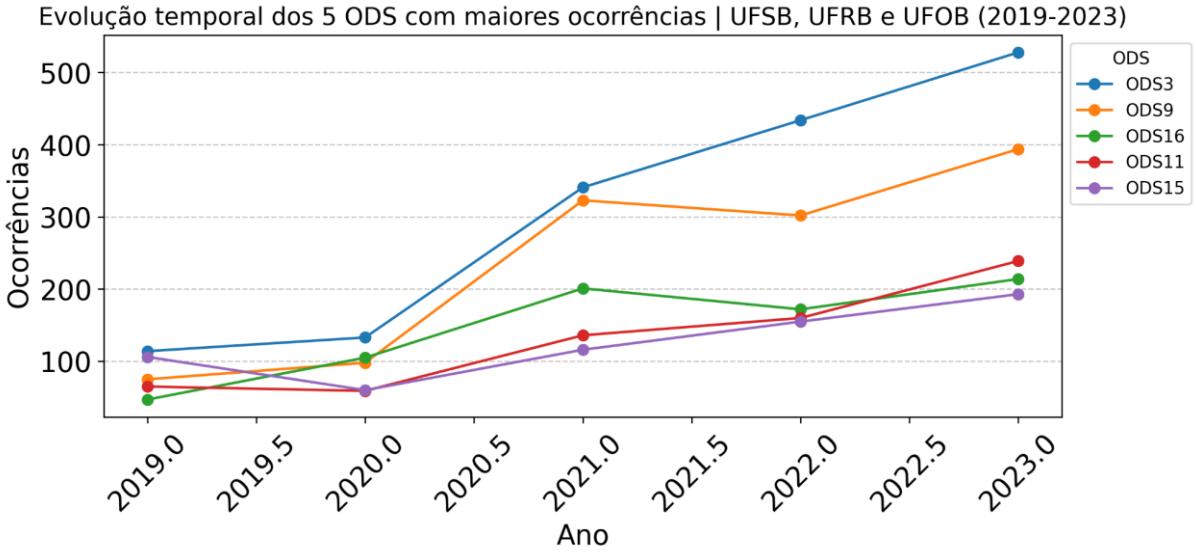

***Figura 16**: Evolução temporal dos 5 ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024)*

O rótulo identificado com maior número de ações (maioria extensionistas) foi o ODS 3, indicando um forte alinhamento das instituições em realizar ações relacionadas com a promoção da saúde e o bem-estar. Algumas iniciativas identificadas foram projetos na área de medicina e saúde pública, e ações voltadas para a promoção de hábitos de vida saudáveis. Houve um crescimento significativo dessas atividades a partir de 2020 (Figura 16), impulsionado pela pandemia de COVID-19 (2020-2022), que intensificou a demanda por ações institucionais de assistência e pesquisa na área da saúde.

A UFRB, que apresentou o maior número de ações alinhadas ao ODS 3, conta com o Centro de Ciências da Saúde (CCS), localizado em Santo Antônio de Jesus, onde abriga cursos de Medicina, Enfermagem, Nutrição e Psicologia. Na UFOB, os cursos de Medicina, Nutrição e Farmácia estão sediados no Centro de Ciências da Saúde em Barreiras. Já a UFSB, conta com curso de Medicina e o Bacharelado Interdisciplinar em Saúde em Teixeira de Freitas.

Por outro lado, o modelo infere algumas áreas que necessitam de atenção. Alguns objetivos, como os ODS 1 (Erradicação da Pobreza), 7 (Energia Limpa e Acessível), 13 (Ação Contra a Mudança Global do Clima) e 17 (Parcerias e Meios de Implementação) apresentaram baixa incidência na análise, o que pode indicar tanto lacunas nas iniciativas quanto desafios na forma como esses temas são descritos nos resumos dos projetos. Vale destacar que, embora o ODS 17 não tenha se sobressaído, é difícil pensar em ações de extensão sem algum tipo de parceria – seja com comunidades, instituições públicas ou outros atores sociais. Isso sugere que a baixa ocorrência pode estar relacionada mais à forma de redação ou à limitação do modelo em captar esse tipo de informação do que à ausência real de articulações colaborativas.

No caso do ODS 13, além da baixa incidência nos dados, o modelo ODSBAHIA-PTBR apresentou métricas menos satisfatórias para esse rótulo em específico. Isso aponta tanto para uma possível escassez de iniciativas diretamente voltadas às mudanças climáticas quanto para a necessidade de aprimorar o desempenho do modelo na identificação desse tema, que muitas vezes aparece de forma transversal ou implícita nos textos. Esses achados reforçam a importância de um olhar atento sobre o que os dados revelam — e, principalmente, sobre o que pode estar ficando invisível —, para que futuras iniciativas acadêmicas e tecnológicas possam preencher essas lacunas.

De modo geral, o modelo inferiu uma diversificação e integração. A **tabela 5** apresenta a coocorrência de ODS entre as ações. Nota-se que um percentual de 79,78% dos projetos das instituições se encontra vinculados a dois ou mais objetivos, indicando uma abordagem holística e multifacetada.

Quantidade de ODS	Um	Dois	Três	Quatro	Cinco ou mais
Total de Projetos	1086	2535	1054	480	216

Tabela 5 - Frequência de Coocorrência dos ODS

A Figura 17 contém a visualização gráfica das coocorrências de multirrótulo dos ODS por meio de um grafo relacional. O tamanho dos nós indica a centralidade de cada ODS, refletindo o número de conexões com outros ODS. A espessura das arestas representa a intensidade das relações, conforme a frequência das coocorrências.

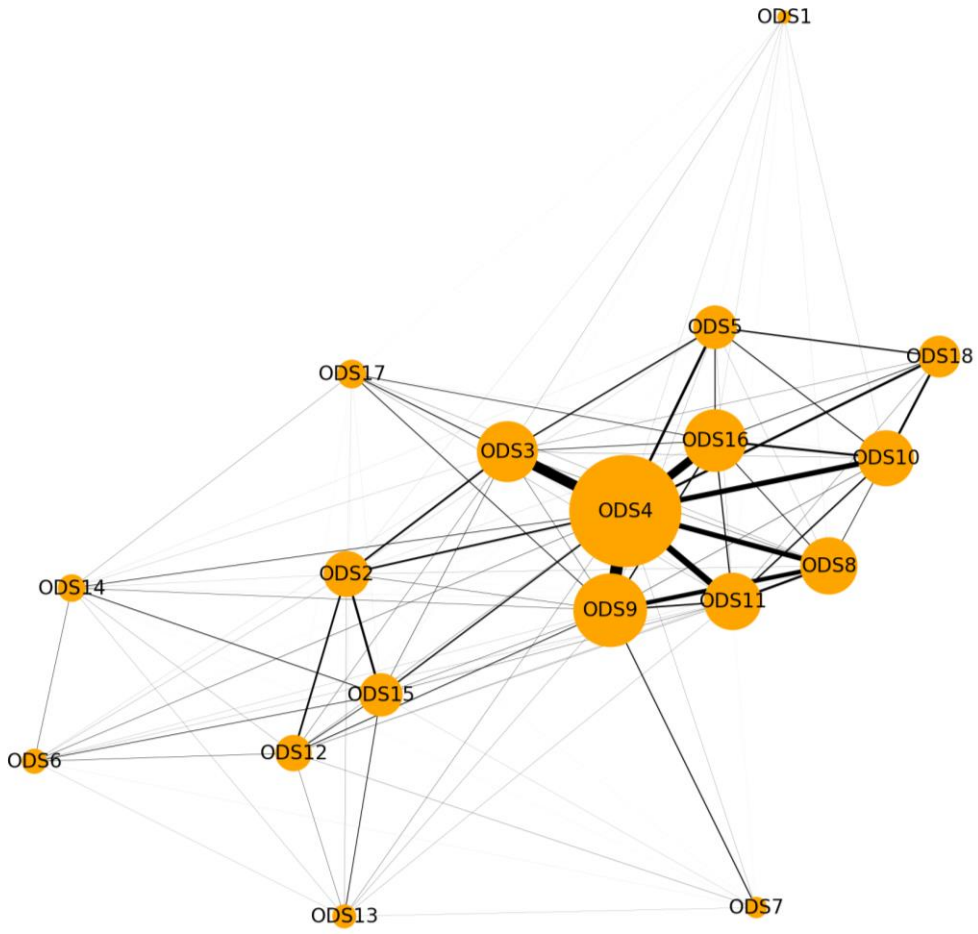

***Figura 17**: Grafo de relações entre os ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024)*

As maiores coocorrências são observadas entre ODS 4 e outros objetivos, com destaque para as relações com os ODS 9, 3, 16 e 11. Essas conexões indicam que muitas dessas ações universitárias estão incorporando temáticas voltadas para saúde, tecnologia e economia. Já o ODS 15 coocorre frequentemente com os ODS 2, 12, 13 e 14, indicando a presença de projetos

focados na conservação ambiental e na agricultura, produção e consumos sustentáveis como parte dos esforços para combater as mudanças climáticas (Figura 18).

O ODS 5, por sua vez, mostra maiores coocorrências com ODS 3, ODS 10 e ODS 18. Esta inter-relação destes objetivos sugere a aplicação de uma abordagem integrada para lidar com desigualdades de gênero em contextos mais amplos de saúde, justiça social e igualdade racial.

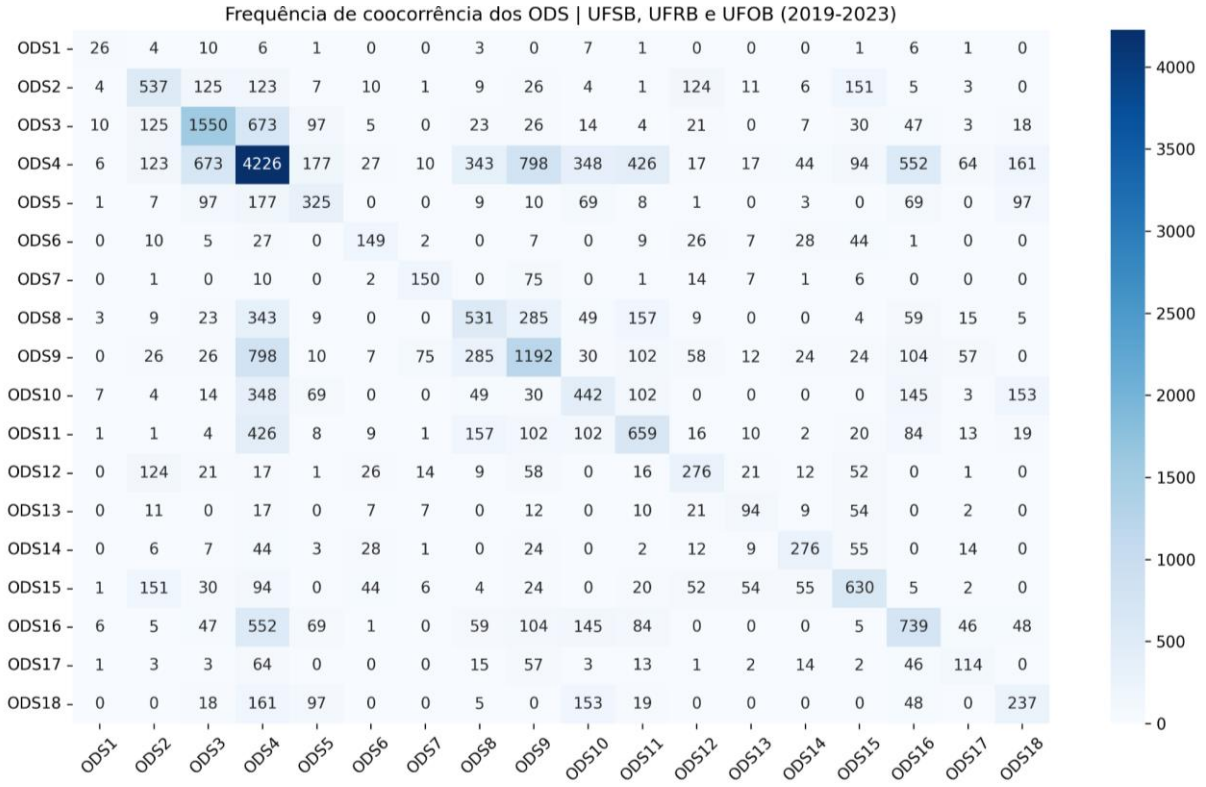

***Figura 18**: Frequência de coocorrências entre os ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024)*

A participação proporcional de cada universidade entre os ODS é apresentada através do mapa de calor presente na Figura 19, a seguir. A UFRB, por possuir o maior número de registros coletados, apresenta uma maior participação nos ODS mapeados, com exceção dos ODS 6, 13, 14 e 15, majoritariamente atendidos pela UFSB. A UFOB possui o menor número de ações coletadas, mas apresenta um equilíbrio entre os ODS, com destaque para os ODS 2, 7 e 12 apresentarem uma participação maior que a UFSB.

Cabe destacar que a análise apresentada na Figura 19 não tem por objetivo estabelecer uma hierarquia entre as universidades ou indicar qual delas realiza mais ou menos ações relacionadas aos ODS. O propósito da visualização é evidenciar como essas ações estão distribuídas proporcionalmente entre os diferentes objetivos e instituições analisadas. Essa

abordagem permite observar possíveis focos temáticos e áreas de concentração de esforços, respeitando as especificidades territoriais, institucionais e operacionais de cada universidade.

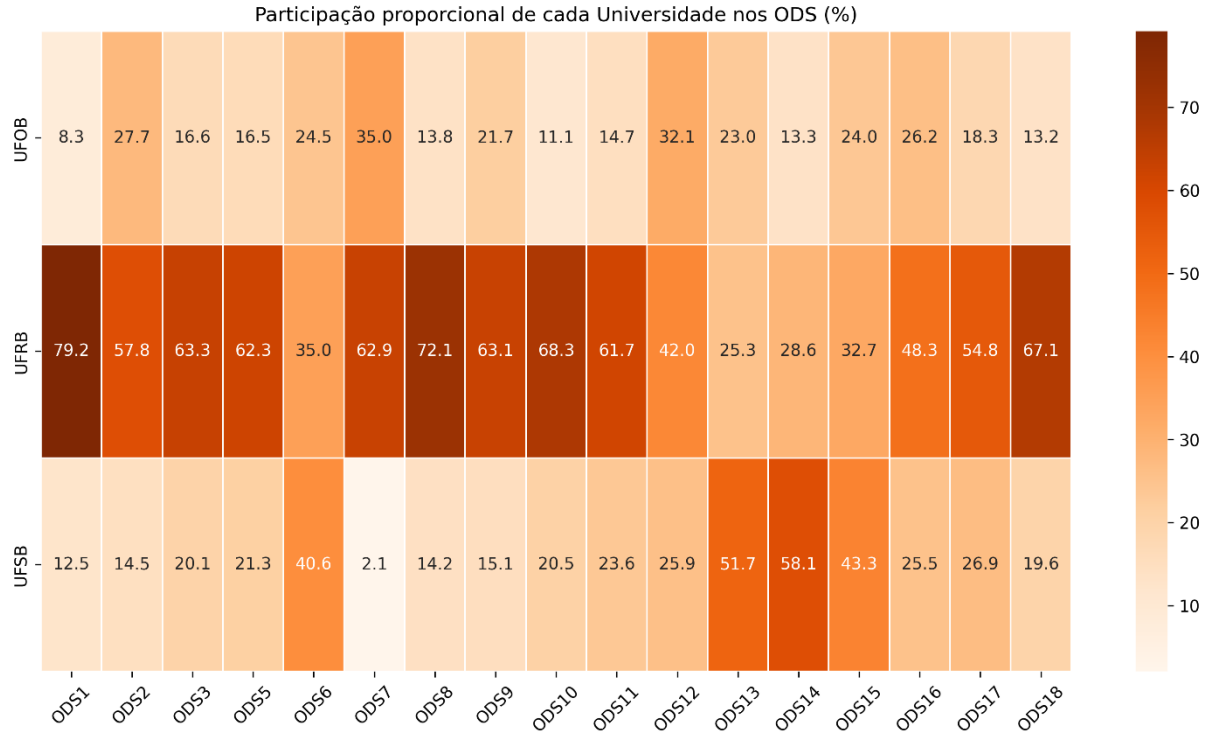

***Figura 19**: Participação proporcional de cada universidade entre os ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023. Autor (2024)*

A figura 20, por sua vez, ilustra a proporção das ações coletadas nas dimensões econômica, social, ambiental, permitindo visualizar o foco e a evolução das ações universitárias em relação a esses pilares do desenvolvimento sustentável entre 2019 e 2023. Para essa análise, os ODS foram agrupados da seguinte forma: Sociedade (ODS 1, 2, 3, 4, 5, 7, 11, 16, 18), Economia (ODS 8, 9, 10, 12) e Ambiental (ODS 6, 13, 14, 15) (ONU, 2015). Entende-se que não é possível avançar em um objetivo sem trabalhar e desenvolver outros de forma conjunta. Sob esta perspectiva, as ações apresentaram uma maior concentração no tripé sociedade (70.46%) que agrupa o maior número de ODS, seguido por economia (20.09%), ambiental (9.45%).

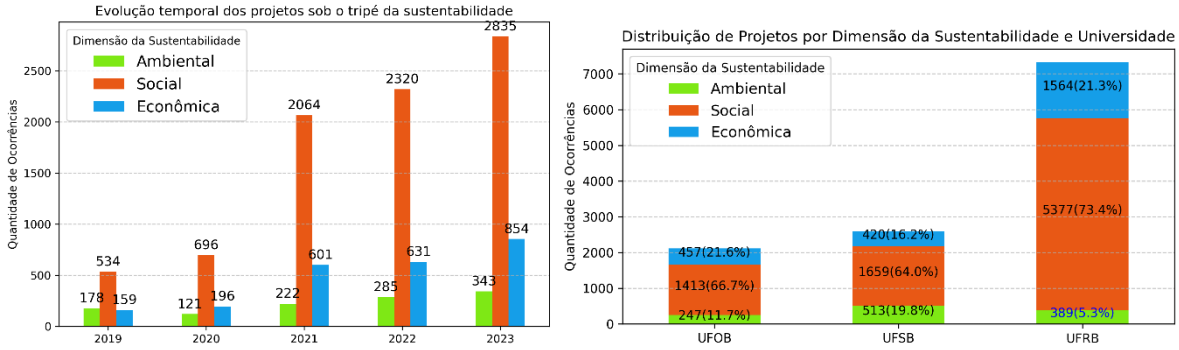

***Figura 20**:Evolução e Distribuição dos ODS nos projetos de pesquisa e extensão coletados do SIGAA das UFSB, UFRB e UFOB no período de 2019 a 2023 sob o Tripé da Sustentabilidade. Autor (2024)*

Também é possível identificar a distribuição das ações, baseado nos 5Ps do desenvolvimento sustentável (Figura 21), enfatizando as cinco áreas cruciais — pessoas (ODS 1 a 6), prosperidade (ODS 7 a 12), planeta (ODS 13 a 15), paz (ODS 16) e parceria (ODS 17), esse agrupamento é baseado em Morton *et al*. (2017).

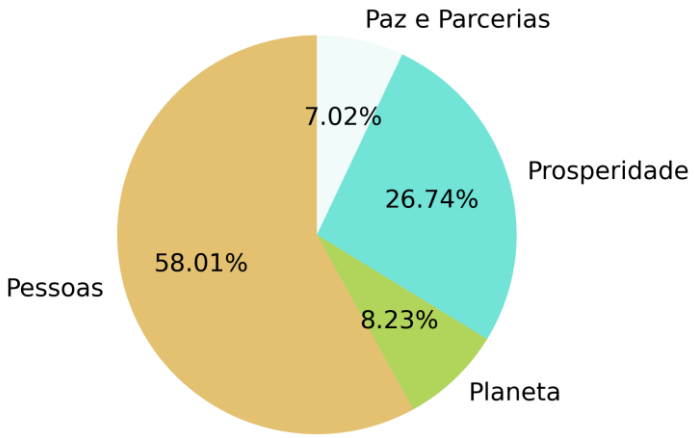

***Figura 21**: Proporção das ações universitárias alinhadas aos 5Ps do desenvolvimento sustentável: Pessoas, Planeta, Prosperidade, Paz e Parcerias. Autor (2024)*

A maioria das ações universitárias analisadas está centrada em objetivos relacionados ao bem-estar humano, como saúde, educação, igualdade de gênero e redução da pobreza (Pessoas, 58.01%), demonstrando um esforço em melhorar a qualidade de vida e atender às necessidades básicas das comunidades. Em seguida, está a promoção da prosperidade econômica (Prosperidade, 26.74%). Isso inclui iniciativas voltadas para trabalho decente, crescimento econômico, inovação e infraestrutura, demonstrando o compromisso com o desenvolvimento econômico sustentável em seus territórios de atuação. As ações relacionadas diretamente ao meio ambiente têm uma representação um pouco menor (Planeta, 8.23%) em comparação com as questões sociais e econômicas. Por fim, as ações voltadas para promover

sociedades pacíficas e inclusivas, bem como fortalecer parcerias para alcançar os ODS representam 7.02% do total.

A distribuição geográfica dos ODS mapeados nas ações universitárias demonstra que cada instituição apresenta um perfil específico de atuação, alinhado tanto com sua estrutura acadêmica quanto com as demandas regionais. (Figura 22). Na cidade de Porto Seguro (UFSB), por exemplo, houve o maior número de registros no ODS 14 em comparação com os demais municípios. Isso indica que a cidade é o principal polo de ações voltadas para a conservação e sustentabilidade dos ecossistemas marinhos e costeiros, o que pode estar relacionado à sua localização litorânea e à presença de iniciativas ambientais focadas na proteção da biodiversidade marinha.

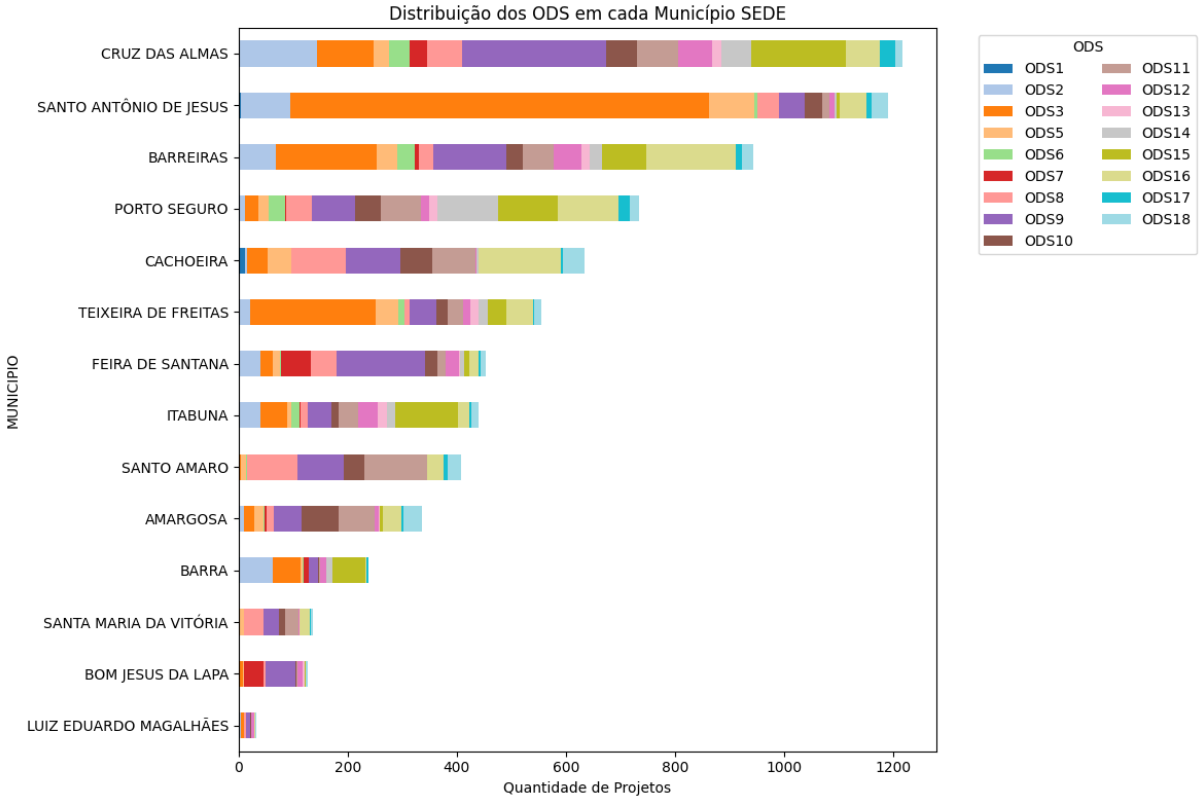

***Figura 22**: Distribuição dos ODS nos municípios sede das UFSB, UFRB e UFOB. Autor (2024)*

Percebe-se que as universidades estão engajadas na busca de incorporar os ODS (Figura 22), o que é uma oportunidade incomparável tanto em termos de ensino quanto de pesquisa, embora Leal Filho *et al.* (2019) considere que "poucas universidades perceberam isso, e muitas estão ficando para trás". Além disso, outras instituições de ensino encontram barreiras que também dificultam a implementação dos ODS, conforme estudado por Serafini *et al*., 2022, que incluem a falta de recursos financeiros, falta de alinhamento entre os conteúdos acadêmicos

e as demandas do território, e a necessidade de melhorar a comunicação interna para acompanhar o progresso do desenvolvimento sustentável.

Uma visualização da distribuição das ações de Pesquisa e Extensão alinhadas aos ODS conforme inferência do modelo ODSBahia-PTBR, separadas por cada instituição é apresentada na Figura 23.

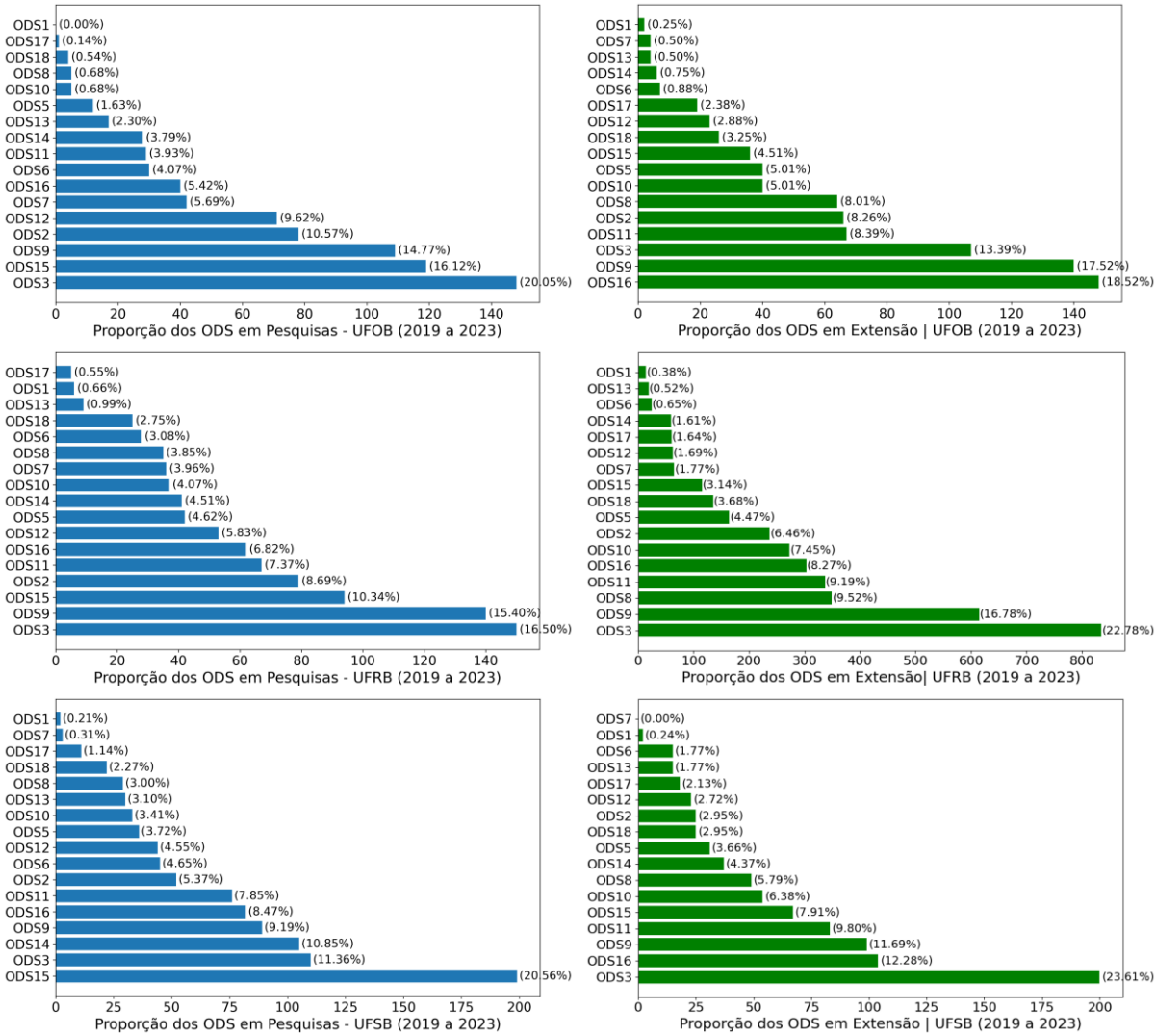

***Figura 23**: Distribuição dos ODS nos Projetos de Pesquisa e Extensão (2019-2023) nas UFSB, UFRB e UFOB. Autor (2024)*

Um aspecto que merece destaque nos resultados diz respeito à atuação da UFSB em relação aos ODS 14 e ODS 15 (Figura 23). Embora o ODS 14 seja, historicamente, um dos menos abordados em escala global nas iniciativas de pesquisa (Shayan *et al*., 2022), ele aparece com frequência significativa nos registros da UFSB, ocupando a terceira posição entre os mais

relacionados. Já o ODS 15 foi o que apresentou o maior número de ocorrências entre as pesquisas da instituição, segundo a inferência realizada pelo modelo. Esse padrão pode ser compreendido a partir da inserção territorial da UFSB, localizada em um território marcado por uma extensa faixa litorânea, ecossistemas marinhos sensíveis e núcleos de biodiversidade da Mata Atlântica. Assim, a universidade reflete em suas ações uma vocação voltada às demandas socioambientais locais, atuando como um agente de interiorização da Agenda 2030 e promovendo o alinhamento entre sua produção acadêmica e os desafios do território onde está inserida.

Por fim, é importante destacar que a contribuição das instituições aqui estudadas inclui, mas não se limitam, apenas aos projetos de pesquisa e extensão. Conforme pontuado por Galvão *et al*. (2024), as universidades públicas brasileiras incorporam os ODS em suas atribuições institucionais por meio de diversas responsabilidades, como a adaptação de normas globais aos contextos locais, a produção de soluções para desafios sociais, ambientais e econômicos e a transferência de conhecimento técnico e científico para a sociedade. Além disso, essas instituições avaliam criticamente os indicadores da Agenda 2030, estabelecendo seus próprios parâmetros para monitoramento e implementação.

A promoção de uma educação de qualidade autêntica, aliada à inclusão social de grupos historicamente marginalizados, fortalece o compromisso das universidades com a democratização do conhecimento e a redução das desigualdades, consolidando seu protagonismo na territorialização dos ODS (Galvão *et al*.,2024).

5.3 Limitações da pesquisa

Importante considerar as limitações desta pesquisa, sobretudo quanto a dependência do SIGAA, uma vez que a análise está limitada aos textos disponíveis na plataforma. Esses registros, por sua natureza, podem não contemplar a totalidade das iniciativas desenvolvidas pelas universidades, tampouco fornecer uma visão aprofundada sobre a efetividade das ações. Além disso, os documentos analisados referem-se apenas aos projetos submetidos, sem comtemplar os relatórios finais de execução, que permitiriam aferir a correspondência entre o que foi planejado e o que efetivamente foi realizado.

Ainda assim, mesmo que os dados estejam baseados nas intenções declaradas nos projetos, eles fornecem uma direção importante sobre os rumos institucionais, prioridades temáticas e áreas de interesse das universidades. Sendo uma fotografia das estratégias propostas e das áreas temáticas que vêm sendo valorizadas no planejamento acadêmico. Esses resultados

não esgotam a complexidade da atuação universitária, mas ampliam as possibilidades de compreensão sobre como essas instituições se articulam com a Agenda 2030 e os ODS, abrindo caminhos para investigações futuras.

Outra limitação relevante refere-se à aplicação do classificador ODSBahia-PTBR. Embora a ferramenta tenha apresentado boas métricas de desempenho em seu desenvolvimento, sua previsão pode ser impactada pela qualidade, clareza e especificidade dos textos analisados. O vocabulário empregado nos resumos das ações analisadas, pode conter termos genéricos ou não fornecer informações detalhadas o suficiente para que o modelo faça uma associação precisa com um ODS específico e, por isso, o modelo pode não ter sido sensível o suficiente para rotulá-los ou ainda as probabilidades de relação desses resumos com um determinado ODS ficaram abaixo do limiar de decisão.

E por fim, os achados são específicos às três instituições analisadas e podem não ser generalizáveis para outras instituições ou contextos. Contudo, a abordagem proposta pode ser adaptada a outros cenários e contribuir para análises comparativas no campo da educação para o desenvolvimento sustentável.

6 CONSIDERAÇÕES FINAIS

Este trabalho explorou duas grandes áreas e como elas podem convergir para promover o desenvolvimento sustentável. De um lado, a Agenda 2030 e os Objetivos de Desenvolvimento Sustentável (ODS), que representam um compromisso global para enfrentar desafios sociais, econômicos e ambientais. Do outro, a Inteligência Artificial (IA), uma tecnologia emergente com potencial para otimizar processos, analisar grandes volumes de dados e apoiar a tomada de decisões estratégicas.

Apresentamos aqui uma abordagem para a classificação de textos em português relacionados aos ODS, através do desenvolvimento e avaliação do modelo ODSBAHIA-PTBR e sua aplicabilidade na análise da contribuição de universidades baianas para a Agenda 2030.

Foi construído e validado um corpus multirrótulo em português, correspondente aos 17 ODS da ONU e três novos ODS propostos no 'Guia Agenda 2030: integrando ODS, Educação e Sociedade'. Em seguida, esse corpus foi usado para ajuste fino nos modelos pré-treinados BERTimbau Base, Albertina e DeBERTinha, sendo eles avaliados e comparados com base nas métricas de precisão, *recall* e *F1-score*.

Com base nos resultados obtidos, houve pequenas variações entre BERTimbau Base, DeBERTinha e Albertina 100M PTBR nas métricas selecionadas. Embora todos esses modelos pré-treinados tenham se mostrado eficazes, o BERTimbau Base foi serviu de base por alcançar um melhor desempenho, sendo então finamente ajustado, dando origem ao ODSBahia-PTBR. Este modelo está publicado em Huggingface (*odsbahia/odsbahia-ptbr*)[8] e disponível abertamente para uso.

Analisamos o ODSBAHIA-PTBR em relação a outros modelos de classificação descritos na literatura que revelou resultados satisfatórios em termos de precisão, *recall* e *F1-Score*. O modelo demonstrou ser capaz de lidar com as particularidades da língua portuguesa, identificando as relações de textos com os ODS com uma precisão de 89%. Esta pesquisa não apenas propôs uma abordagem objetiva para a classificação de textos em português em relação aos ODS, mas também demonstrou a importância da escolha e adaptação adequada de modelos pré-treinados para diferentes contextos de aplicação.

[8] Disponível em: https://huggingface.co/odsbahia/odsbahia-ptbr

Além disso, este trabalho analisou a contribuição de universidades baianas para a implementação dos ODS, buscando identificar como elas fazem o que falam sobre a Agenda 2030.

Os objetivos desta pesquisa foram alcançados ao mapear e categorizar as ações de pesquisa e extensão das universidades UFSB, UFRB e UFOB em relação aos ODS, por meio da aplicação de técnicas de Processamento de Linguagem Natural (PLN) utilizando o ODSBAHIA-PTBR na análise textual dos resumos dos projetos coletados. A análise permitiu identificar as iniciativas desenvolvidas entre 2019 e 2023, evidenciando as prioridades institucionais e a distribuição geográfica dos ODS nos municípios sede dessas instituições. Os resultados destacaram padrões de concentração e lacunas na implementação dos ODS nas regiões estudadas, contribuindo para uma compreensão das ações universitárias em relação a Agenda 2030

A análise revelou uma forte concentração de projetos voltados para os ODS 3, 9 e 16, refletindo o envolvimento das universidades na promoção da saúde, inovação tecnológica e governança. Por outro lado, objetivos como ODS 1, 7 e 13 apresentaram baixa incidência, sugerindo oportunidades para futuras iniciativas acadêmicas nessas áreas.

O ODS 4 (Educação de Qualidade) se destacou como eixo central na interconexão com outros ODS. Essa forte coocorrência com outros ODS evidencia a educação como um vetor transversal para o desenvolvimento sustentável e que muitas das ações realizadas nos últimos anos pelas instituições aqui analisadas têm abordagem multidisciplinar, conectando ensino, pesquisa e extensão a temas estratégicos como saúde, inovação, tecnologia e desenvolvimento econômico.

É importante destacar que os resumos dos projetos analisados oferecem apenas uma visão resumida das iniciativas, destacando suas características principais e ideia geral da ação. Isso significa que o modelo ODSBAHIA-PTBR conseguiu captar o contexto e as intenções descritas, mas não tem como avaliar o processo detalhado ou os resultados alcançados por elas.

Ademais, os resultados obtidos demonstram que a aplicação de ferramentas de IA na análise dos ODS pode ampliar a capacidade de monitoramento e avaliação do impacto das iniciativas relacionadas ao desenvolvimento sustentável. Nesta pesquisa, por exemplo, foi possível identificar padrões, destacar áreas prioritárias e como instituições de ensino superior (IES) contribuem para os diferentes ODS.

Em síntese, os achados desta pesquisa reforçam o papel estratégico das universidades na concretização da Agenda 2030, ao evidenciar como suas ações se alinham aos desafios

globais de desenvolvimento. A incorporação de tecnologias de Inteligência Artificial nesse processo, como demonstrado com o uso do ODSBAHIA-PTBR, amplia as possibilidades de análise, planejamento e tomada de decisão, fortalecendo a atuação das IES como agentes transformadores. Essa articulação entre ciência, tecnologia e políticas públicas representa um caminho promissor para transformar dados em conhecimento e, sobretudo, em impacto positivo na sociedade.

Mais do que a categorização em 17, 18 ou 20 Objetivos de Desenvolvimento Sustentável, é fundamental reconhecer a relevância dos temas que estão na base dessas agendas. Questões como equidade, justiça social, enfrentamento das mudanças climáticas, promoção da saúde, educação de qualidade e inclusão não se encerram com ciclos políticos ou marcos temporais como 2030. Esses desafios são estruturais e continuarão sendo urgentes, independentemente da existência formal dos ODS ou de como forem reorganizados em futuras agendas. A preocupação com os princípios que orientam a Agenda 2030 precisa seguir no centro das ações institucionais, acadêmicas e políticas. Tecnologias como a IA podem fortalecer esse compromisso, ao ampliar nossa capacidade de gerar conhecimento, identificar lacunas e propor caminhos mais justos e sustentáveis.

Como desdobramento deste trabalho, pesquisas futuras podem explorar diferentes caminhos para aprimorar o modelo ODSBAHIA-PTBR. Entre as possibilidades, destaca-se a otimização do desempenho em rótulos específicos dos ODS, a ampliação e diversificação do corpus de treinamento, bem como a integração de novas fontes de dados para validação.

Outra frente consiste na adaptação e aplicação da abordagem proposta aqui em outras instituições e/ou contextos, o que permitirá avaliar a generalização e possíveis ajustes metodológicos. Também se recomenda o desenvolvimento de modelos próprios de classificação, mais alinhados às particularidades regionais ou institucionais, com vistas a análises mais refinadas e contextualizadas.

Além disso, é fundamental investigar as lacunas na pesquisa e na implementação de ações em determinados ODS que merecem uma atenção mais aprofundada. Alguns dos ODS, apesar de sua importância, ainda não receberam a devida consideração necessária para garantir que a Agenda 2030 seja verdadeiramente inclusiva e que ninguém seja deixado para trás.

REFERÊNCIAS

ABRAMOVAY R. Muito além da economia verde. 1ª. ed. São Paulo: Editora Abril, v. 01. 2012. 247p.

ADHIKARI, A. et al. DocBERT: BERT for Document Classification. **arXiv (Cornell University)**, 1 jan. 2019. DOI: https://doi.org/10.48550/arXiv.1904.08398.

AIBA, Anuário Região Oeste da Bahia: safra/crop 2022/2023. Barreiras, Editora Gazeta Santa Cruz, 2023. Disponível em <https://aiba.org.br/arquivos/anuarios/AIBA_ANU%C3%81RIO%20FINAL%202022.23.pdf>.Acesso em 22 nov 2024.

AIME, Artificial intelligence in Medical Epidemiology. Disponível em: <https://www.nesta.org.uk/feature/collective-crisis-intelligence-case-studies/artificial-intelligence-medical-epidemiology-aime/>. Acesso em: 6 set. 2024.

ALCÁNTARA-RUBIO, L. *et al*., The implementation of the SDGs in universities: a systematic review. Environmental Education Research, p. 1–31, 2022.

ALLEN, C. *et al*., Priorities for science to support national implementation of the sustainable development goals: A review of progress and gaps, Sustainable Development, 2021.

AMAGLOBELI, D. *et al*. The future of saving: the role of pension system design in an aging world. Washington, D.C.: International Monetary Fund, 2019. (IMF Staff Discussion Note, SDN/19/01). Disponível em: https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2019/01/09/The-Future-of-Saving-The-Role-of-Pension-System-Design-in-an-Aging-World-45138. Acesso em: 28 dez. 2025.

BAHDANAU, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. **arXiv (Cornell University)**, 1 set. 2014. DOI: https://doi.org/10.48550/arXiv.1409.0473

BANCO MUNDIAL. Relatório Anual de 2019 do Banco Mundial, Washington, DC: Banco Mundial. 2019. DOI: 10.1596/978-1-4648-1476-1.

BEDIN, É. P.; FARIA, L.C. Sustentabilidade em instituições de ensino superior: revisão sistemática, Revista Brasileira de Gestão Ambiental e Sustentabilidade, v. 8, n. 19, p. 1183–1201, 2021.

BELTAGY, I., Lo, K. & Cohan, A. SciBERT: A pretrained language model for scientific text. **arXiv (Cornell University)**, 26 mar 2019. DOI: https://doi.org/10.48550/arXiv.1903.10676.

BELTAGY, I., Peters, M. E. & Cohan, A. Longformer: The Long-Document Transformer. **arXiv (Cornell University)**, 10 abr. 2020. DOI: https://doi.org/10.48550/arXiv.2004.05150.

BHATTACHARJEE, N. V. *et al*. Global fertility in 204 countries and territories, 1950–2021, with forecasts to 2100: a comprehensive demographic analysis for the Global Burden of

Disease Study 2021. The Lancet, v. 403, n. 10440, p. 2057-2099, 2023. DOI: https://doi.org/10.1016/S0140-6736(24)00550-6

BOJANOWSKI, P. *et al*. Enriching word vectors with subword information. In: Proceedings of the 2017 Conference of the Association for Computational Linguistics (ACL). Vancouver: Association for Computational Linguistics, 2017. p. 135–142. DOI: https://doi.org/10.18653/v1/P17-1010

BOMMASANI, R. *et al*. On the opportunities and risks of foundation models. **arXiv (Cornell University)**, 16 ago. 2021. DOI: https://doi.org/10.48550/arXiv.2108.07258.

BORSATTO, J. M. L. S. *et al*. O uso de inteligência artificial para mapeamento das ações de extensão alinhadas aos ODS: um estudo na Universidade Federal de Uberlândia. In: ENCONTRO NACIONAL SOBRE GESTÃO EMPRESARIAL E MEIO AMBIENTE, 24., 2022, São Paulo. Anais [...]. São Paulo: ENGEMA, 2022. Disponível em: https://engemausp.submissao.com.br/24/anais/arquivos/138.pdf?v=1738779329. Acesso em: 5 jan. 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovação. MCTI elabora retrato mais recente da ciência, tecnologia e inovação no Brasil. Brasília, 05 maio 2023. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2023/05/mcti-elabora-retrato-mais-recente-da-ciencia-tecnologia-e-inovacao-no-brasil. Acesso em: 20 de jan. 2025.

BRASIL. Ministério da Educação. Universidades Federais na Bahia. Disponível em: https://www.gov.br/mec/pt-br/areas-de-atuacao/es/universidades-federais/bahia. Acesso em: 22 nov. 2024.

BRASIL. Relatório nacional voluntário, Brasil 2024. Brasília: Presidência da República, Secretaria-Geral da Presidência da República, 2024. 343 p.: il. ISBN 978-65-86360-45-5.

BROWN, T. B. *et al*. Language Models are Few-Shot Learners. **arXiv (Cornell University)**, 26 mai. 2020. DOI: https://doi.org/10.48550/arXiv.2005.14165.

CABRAL, R.; GEHRE, T. Guia agenda 2030: Integrando ODS, Educação e Sociedade. 1 Ed. São Paulo, 2020. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/60bba95b-fe49-40dd-b01b-7adc68e961a0/content>. Acesso em: 20 jan. 2024.

CALLAGHAN, M. *et al*. (2022). Machine learning-based evidence and attribution mapping of 100,000 climate impact studies, 22 jun. 2022, PREPRINT (Version 2) available at Research Square [https://doi.org/10.21203/rs.3.rs-783398/v2]

CAMPIOTTI, I. *et al*. DeBERTinha: A multistep approach to adapt DebertaV3 XSmall for Brazilian Portuguese natural language processing task. **arXiv (Cornell University)**, 28 set. 2023. DOI: https://doi.org/10.48550/arXiv.2309.16844

CASTRO, C.M. A Prática da Pesquisa. 2a. Ed. Pearson Prentice Hall. (2o ed.). São Paulo, Pearson Prentice Hall, 156 p, 2006.

CHANKSELIANI, M.; MCCOWAN, T. Higher education and the Sustainable Development Goals. Higher Education, v. 81, 2020. DOI: https://doi.org/10.1007/s10734-020-00652-w

CHENG, L. *et al*. Record High Temperatures in the Ocean in 2024. Advances in Atmospheric Sciences, 10 jan. 2025. DOI: https://link.springer.com/article/10.1007/s00376-025-4541-3

Cho, K. *et al*. Learning phrase representations using RNN encoder–decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Anais...Stroudsburg, PA, USA: Association for Computational Linguistics. DOI: https://doi.org/10.3115/v1/D14-1179

CMMAD, Comissão Mundial sobre Meio Ambiente e Desenvolvimento. Nosso futuro comum. Rio de Janeiro: Editora da Fundação Getulio Vargas, 1991.

COSTA E SILVA, M. da V. *et al*. O Papel das universidades na institucionalização da agenda 2030: Análise da implementação dos Objetivos de Desenvolvimento Sustentável nas universidades federais da Amazônia Legal, BRASIL. Capim Dourado: Diálogos em Extensão, [S. l.], v. 6, n. 2, p. 44–77, 2024. DOI: https://doi.org/10.20873/v6/A.

Da Costa, P. *et al*. BERTabaporu: Assessing a genre-specific language model for Portuguese NLP. Proceedings of the Conference Recent Advances in Natural Language Processing - Large Language Models for Natural Language Processings. Anais...INCOMA Ltd., Shoumen, BULGARIA.

DALY H. E. Towards an environmental economics. Land Economics, 1991, 67(2): 255-259.

DALY H. E. Uneconomic growth: in theory, in fact, in history, and its relation to globalization. Clemens Lectures Series, Saint's John University, 1999.

DALY H.E. "Economics for a Full World", Great Transition Initiative, 2015.

DE IORIO, S. *et al*., Determinant Factors of SDG Disclosure in the University Context, Administrative Sciences, v. 12, n. 1, p. 21, 2022.

DEMIRCI, M. *et al*. Modeling of groundwater level using artificial intelligence techniques: A case study of reyhanli region in Turkey. Applied Ecology and Environmental Research, 17 (2) (2019), pp. 2651-2663. DOI: http://dx.doi.org/10.15666/aeer/1702_26512663

DENNY, M. J.; SPIRLING, A. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. Political Analysis, Cambridge University Press, v. 26, n. 2, p. 168–189, abr. 2018. ISSN 1047-1987, 1476-4989. Disponível em: http://dx.doi.org/10.2139/ssrn.2849145.

DEUTSCHLAND.DE. A contribuição da Alemanha para a proteção climática. 2024. Disponível em: https://www.deutschland.de/pt-br/topic/meio-ambiente/objetivos-climaticos-alemanha-transicao-energetica-protecao-climatica. Acesso em: 18 dez. 2024.

DEVLIN, J. *et al*. BERT: Pre-training of deep bidirectional Transformers for language understanding. **arXiv (Cornell University)**, 11 out. 2018. DOI: https://doi.org10.48550/arXiv.1810.04805

FACELI, K. *et al*. Inteligência artificial: uma abordagem de aprendizado de máquina. 2. ed. 4. reimpr. Rio de Janeiro: LTC, 2021.

FAN, A. *et al*. Beyond English-Centric multilingual machine translation. **arXiv (Cornell University)**, 21 out. 2020. DOI: https://doi.org/10.48550/arXiv.2010.11125

FAVARETO, A. A Agenda 2030, a Dimensão Territorial do Desenvolvimento Brasileiro e os Motores da Transição para a Sustentabilidade. Novos Caminhos do Desenvolvimento, 2020. DOI: https://doi.org/10.1007/978-3-030-56096-6_10.

FIA, M. *et al*. How Higher Education Institutions Walk Their Talk on the 2030 Agenda: A Systematic Literature Review. Higher Education Policy, 2022. DOI: https://doi.org/10.1057/s41307-022-00277-x

FROIO, L.R.; GONÇALVES, P.T.C. The Power of Localities: Subnational Governments and the Territorialization of SDGs. In: Galvao, T.G., de Menezes, H.Z. (eds) The Quest for the Sustainable Development Goals. Sustainable Development Goals Series. Springer, Cham, p. 45–57, 23 ago. 2024. DOI: https://doi.org/10.1007/978-3-031-59279-9_4

FUX, L. *et al*. Classificação de processos judiciais segundo Objetivos de Desenvolvimento Sustentável da Agenda ONU 2030. 2022. Revista da CGU, v. 14, n. 26.

GALVÃO, T. G. *et al*. Challenges and Opportunities of a Brazilian Network of Universities for the SDG Implementation. In: Galvao, T.G., de Menezes, H.Z. (eds) The Quest for the Sustainable Development Goals. Sustainable Development Goals Series. Springer, Cham, p. 59–69, 23 ago. 2024. DOI: https://doi.org/10.1007/978-3-031-59279-9_5

GALVÃO, T. G.; FREITAS, R.; MARTINS, S. L. Building an Ecosystem of Innovative SDG-Led Solutions: Brazilian Civil Society Organizations Driving the 2030 Agenda Implementation. In: Galvao, T.G., de Menezes, H.Z. (eds) The Quest for the Sustainable Development Goals. Sustainable Development Goals Series. Springer, Cham, p. 31–44, 23 ago. 2024. DOI: https://doi.org/10.1007/978-3-031-59279-9_3

GANEM, R. S. De Estocolmo à Rio+20: Avanço ou Retrocesso? Cadernos Aslegis, n.45, p. 31-62, jan/abr 2012. Disponível em: <http://bd.camara.leg.br/bd/handle/bdcamara/12297>. Acesso em: 16 dez. 2023

GARDNER, H. Inteligências múltiplas - a teoria na prática. Porto Alegre: Artes Médicas, 1995.

GIL, Antônio Carlos. Métodos e técnicas de pesquisa social. 6a ed. São Paulo. Atlas. 2009.

GOLDBERG, Y. Neural Network Methods in Natural Language Processing. San Rafael, CA, USA.: Morgan & Claypool, 2017. (Synthesis Lectures on Human Language Technologies). ISBN 978-1-62705-298-6.

GOTTFREDSON, L. S. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. Intelligence, v. 24, n. 1, p. 13-23, 1997.

GRUPTA, K. A Study of Human Development Achievements in Norway. International Journal for Multidisciplinary Research (IJFMR), vol. 6, no. 6, Nov.-Dec. 2024, DOI: 10.36948/ijfmr.2024.v06i06.32410.

GUISIANO, J. E., Chiky, R. & Mello, J. de. SDG-Meter: a deep learning based tool for automatic text classification of the Sustainable Development Goals. hal.science. Disponível em: <>. Acesso em: 12 abr. 2024.

HAJIKHANI, A. & Suominen, A. Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection. 2022. Scientometrics, v. 127, n. 11, p. 6661–6693. DOI: https://doi.org/10.1007/s11192-022-04358-x

HAYKIN, S. Redes Neurais: Princípios e Prática. Porto Alegre: Grupo A, 2007. E-book. ISBN 9788577800865. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9788577800865/. Acesso em: 04 ago. 2024.

HE, P. *et al*. DeBERTa: Decoding-enhanced BERT with disentangled attention. **arXiv (Cornell University)**, 5 jun. 2020. DOI: https://doi.org/10.48550/arXiv.2006.03654.

HIRSCHBERG, J.; MANNING, C. D. Advances in natural language processing. Science, American Association for the Advancement of Science, v. 349, n. 6245, p. 261–266, jul. 2015. Disponível em: https://www.science.org/doi/full/10.1126/science.aaa8685.

HOCHREITER, S.; SCHMIDHUBER, J.. “Long short-term memory”. Neural Computation, v. 9, n. 8, p. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735

HOSSIN; S. A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 2020, v. 5, n. 2, p. 01–11.

HU, D. An introductory survey on attention mechanisms in NLP problems. Em: Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2020. p. 432–448. DOI: https://doi.org/10.1007/978-3-030-29513-4_31.

HUGGING FACE. Hugging Face – On a mission to solve NLP, one commit at a time. huggingface.co. Disponível em: <https://huggingface.co/>. Acesso em: 11 dez. 2023.

HUGGING FACE. Hugging Face – On a mission to solve NLP, one commit at a time. huggingface.co. Disponível em: <https://huggingface.co/odsbahia/odsbahia-ptbr>. Acesso em: 11 dez. 2024.

IBGE - Instituto Brasileiro de Geografia e Estatística. Plataforma ODS Brasil: Indicadores para os Objetivos de Desenvolvimento Sustentável. Disponível em: https://odsbrasil.gov.br/. Acesso em: 28 de set 2024.

INÁCIO, M. A. F. *et al*. O desempenho das universidades brasileiras no alcance dos Objetivos de Desenvolvimento Sustentável: Uma análise do período de 2020 a 2024. In: SEMINÁRIO DE ADMINISTRAÇÃO – SEMEAD, 27., 2024, São Paulo. Anais [...]. São

Paulo: FEA-USP, 2022. Disponível em: https://login.semead.com.br/27semead/anais/arquivos/2050.pdf. Acesso em: 8 jan. 2025

IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K.

IPEA, Instituto de Pesquisa Econômica Aplicada. Objetivos de Desenvolvimento do Milênio – Relatório nacional de acompanhamento. Brasília, 2004. Disponível em: <https://www.gov.br/abc/pt-br/centrais-de-conteudo/publicacoes/documentos/1_relatorionacionalacompanhamentoodm.pdf>. Acesso em 24 ago 2024.

ISLEK, I. & Oguducu, S. G. (2020). A hybrid recommendation system based on bidirectional encoder representations. Em: ECML PKDD 2020 Workshops. Cham: Springer International Publishing, 2020. p. 225–236. DOI: https://doi.org/10.1007/978-3-030-65965-3_14.

IZBICKI, R.; SANTOS, T. M. dos. Aprendizado de máquina: uma abordagem estatística [livro eletrônico]. São Carlos, SP: Rafael Izbicki, 2020. PDF. ISBN 978-65-00-02410-4.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. Electronic Markets, v. 31, n. 3, p. 685–695, 9 2021. DOI: https://doi.org/10.1007/s12525-021-00475-2

JARVIE, M. E. Brundtland Report. Encyclopedia Britannica , 2016. Disponível em: <https://www.britannica.com/topic/Brundtland-Report>. Acesso em 9 ago 2024.

JESUS, G.S., Gestal, P.F., Silva, A.F. *et al*. Effects of conditional cash transfers on tuberculosis incidence and mortality according to race, ethnicity and socioeconomic factors in the 100 Million Brazilian Cohort. Nat Med (2025). https://doi.org/10.1038/s41591-024-03381-0

KALIYAR, R. K., Goswami, A. & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia tools and applications, v. 80, n. 8, p. 11765–11788. DOI: https://doi.org/10.1007/s11042-020-10183-2.

KHAN, J. *et al*. A comprehensive review of conventional, machine leaning, and deep learning models for groundwater level (GWL) forecasting. Applied sciences (Basel, Switzerland), v. 13, n. 4, p. 2743, 2023. DOI:

KHARLASHKIN, L. *et al*. Predicting sustainable development goals using course descriptions -- from LLMs to conventional foundation models. **arXiv (Cornell University)**, 26 fev. 2024. DOI: https://doi.org/10.48550/arXiv.2402.16420.

KHURANA, D. *et al*. Natural language processing: state of the art, current trends and challenges. 2023. Multimedia tools and applications, v. 82, n. 3, p. 3713–3744. https://doi.org/10.1007/s11042-022-13428-4.

LAGO, A. A. C. do. Estocolmo, Rio, Joanesburgo: o Brasil e as três conferências ambientais das Nações Unidas. Brasília: Instituto Rio Branco (IRBr); Fundação Alexandre de Gusmão (FUNAG); Ministério das Relações Exteriores, 2006. Disponível em: <http://funag.gov.br/loja/download/903-Estocolmo_Rio_Joanesburgo.pdf>. Acesso em 08 jun 2023

LEAL FILHO, W. *et al*, Sustainable Development Goals and sustainability teaching at universities: Falling behind or getting ahead of the pack? Journal of Cleaner Production, v. 232, p. 285–294, 2019.

LEAL FILHO, W. *et al*., A framework for the implementation of the Sustainable Development Goals in university programmes. Journal of Cleaner Production, v. 299, p. 126915, 2021. Disponível em: <https://doi.org/10.1016/j.jclepro.2021.126915>. Acesso em: 22 mar. 2023.

LECUN, Y. *et al*. "Backpropagation Applied to Handwritten Zip Code Recognition," in Neural Computation, vol. 1, no. 4, pp. 541-551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.

LEE, J. *et al*. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. 2020. Bioinformatics (Oxford, England), v. 36, n. 4, p. 1234–1240. DOI: https://doi.org/10.1093/bioinformatics/btz68.

LEFRANÇOIS, G. R. Teorias da Aprendizagem – tradução da 5ª ed. São Paulo: Cengage Learning, 2013.

LIU, Y. *et al*. (2019). RoBERTa: A robustly optimized BERT pretraining approach. DOI: .

LUONG, M.-T., Pham, H. & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. DOI: https://doi.org10.48550/arXiv.1508.04025.

MANNING, C., Raghavan, P. & Schutze, H. (2008). Introduction to Information Retrieval. [s.l.]: Cambridge University Press. ISBN 978-0-521-86571-5.

MARINHO M. B. Universidades e sustentabilidade. Uma pesquisa em Instituições de Educação Superior Brasileiras. Tese de Doutorado. Programa de Engenharia Industrial-PEI. Universidade Federal da Bahia. Escola Politécnica. 2014. 181pp.

MARTINS, B. A., & Rückert, F. Q. O Programa Bolsa Família e a educação: uma análise da produção científica brasileira fundamentada na base de dados SciELO (2003-2018). Revista Brasileira De Educação, 24, e240061, 2019. ttps://doi.org/10.1590/S1413-24782019240061

MATSUI, T. *et al*. A natural language processing model for supporting sustainable development goals: translating semantics, visualizing nexus, and connecting stakeholders. Sustainability science, v. 17, n. 3, p. 969–985, 2022. DOI: https://doi.org/10.1007/s11625-022-01093-3.

MAXWELL. Aprendizado por reforço. [S.l.]: Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio, 2017. Disponível em: <https://www.maxwell.vrac.puc-rio.br/19637/19637_4.PDF> Acesso em 13 dez. 2023.

MCARTHUR, J. W.; RASMUSSEN, K. Change of pace: Accelerations and advances during the Millennium Development Goal era. World Development, v. 105, p. 132–143, maio 2018.

MCCARTHY, J. What is artificial intelligence? - Stanford University. Computer Science Department, 2007. Disponível em: <http://www-formal.stanford.edu/jmc/whatisai/> Acesso em: 23 jan 2024.

McCULLOCH, W. S; PITTS, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, v. 5, n. 4, p. 115-133, 1943. DOI: https://doi.org/10.1007/BF02478259

MEADOWS, D. L., MEADOWS, D. H., RANDERS, J. & BEHRENS, W.W. Limites do crescimento - um relatório para o Projeto do Clube de Roma sobre o dilema da humanidade. São Paulo: Ed. Perspectiva, 1972.

MEDEIROS, L. F. de. Inteligência Artificial Aplicada: Uma Abordagem Introdutória. Curitiba: InterSaberes, 2018.

MEDEIROS, L. F. Redes Neurais em Delphi. 2. ed. Florianópolis: Visual Books, 2007.

MENEZES, M; A. de. A Inteligência Artificial versus a Inteligência Humana. Saber Humano: Revista Científica da Faculdade Antonio Meneghetti, v. 13, n. 22, p. 220–239, 2023. DOI: https://doi.org/10.18815/sh.2023v13n22.576

MERIZALDE, Y. *et al*. Maintenance models applied to wind turbines. A comprehensive overview. Energies, v. 12, n. 2, p. 225, 2019. DOI: https://doi.org/10.3390/en12020225

MIKOLOV, T. *et al*. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

MINSKY, M. A Sociedade da Mente. São Paulo: Perspectiva, 1985.

MIRANDA, K. F.; Amaral Neto, M., Sousa, R. da P., & Coelho, R. de F.. (2020). Sustainable Forest Management in Community Use Protected Areas in the Amazon. Sociedade & Natureza, 32, 799–814. DOI: https://doi.org/10.14393/SN-v32-2020-51621

MORTON, S. *et al*. Sustainable Development Goals (SDGs), and their implementation. British Medical Bulletin, v. 124, n. 1, p. 81–90, 2017. Disponível em: <https://doi.org/10.1093/bmb/ldx031>. Acesso em: 19 maio 2023.

MOUSTAFA, Mona. (2022). On SDG 18: War legacy, Resilience, and Healing in Uncertain Times! United Nations Development Programme, Lao PDR, 8 dez. 2022. Disponível em: <https://www.undp.org/laopdr/blog/sdg-18-war-legacy-resilience-and-healing-uncertain-times>. Acesso em: 19 jun. 2024.

NGUYEN, D. Q., Vu, T., & Nguyen, A. T. (2020). BERTweet: A pre-trained language model for English Tweets. DOI: https://doi.org/10.48550/arXiv.2005.10200.

NISHANT, R.; KENNEDY, M.; CORBETT, J. Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. International journal of information management, v. 53, n. 102104, p. 102104, 2020. DOI:

ONU (2019). World Population Prospects 2019: Highlights. Disponível em: https://population.un.org/wpp/Publications/Files/WPP2019_PressRelease_EN.pdf Acesso em 01 de agosto de 2020.

ONU (United Nations) 2000. Declaração do Milénio das Nações Unidas. Disponível em: <https://www.undp.org/sites/g/files/zskgke326/files/migration/br/undp-br-declaracao_do_milenio.pdf> Acesso em: 24 de abril de 2024.

ONU (United Nations) 2008. Universal Declaration of Human Rights. Escrito por TRINDADE, A. A. C. Disponível em: https://legal.un.org/avl/pdf/ha/udhr/udhr_e.pdf. Acesso em: 27 ago. 2024.

ONU (United Nations) 2013. ONU avalia progresso dos Objetivos do Milênio e pede esforço pelo cumprimento das metas até 2015. Disponível em: <https://brasil.un.org/pt-br/63467-onu-avalia-progresso-dos-objetivos-do-mil%C3%AAnio-e-pede-esfor%C3%A7o-pelo-cumprimento-das-metas-at%C3%A9>. Acesso em 24 ago 2024.

ONU (United Nations) 2015. Transformando Nosso Mundo: A Agenda 2030 para o Desenvolvimento Sustentável. Disponível em: <https://brasil.un.org/sites/default/files/2020-09/agenda2030-pt-br.pdf> Acesso em: 24 de abril de 2024.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Declaração do Milênio. 2000. Disponível em: https://www.undp.org/pt/brazil/publications/declaracao-do-milenio. Acesso em: 28 ago. 2024.

OSES-ALC- Observatório da Sustentabilidade do Ensino Superior na América Latina e no Caribe. Avanços no conhecimento, compromisso e contribuições para os Objetivos de Desenvolvimento Sustentável nas instituições de ensino superior da América Latina e do Caribe – Relatório IES brasileiras. [S.l.: s.n.], 2024. Disponível em: https://oses-alc.net/wp-content/uploads/2024/01/2024-02-22-OSES-ALC-Informe-Brasil.pdf. Acesso em: 5 jan. 2025.

OTTER, D. W., MEDINA, J. R. & KALITA, J. K. (2021) A survey of the usages of deep learning for natural language processing. IEEE transactions on neural networks and learning systems, v. 32, n. 2, p. 604–624. DOI: https://doi.org/10.1109/TNNLS.2020.2979670

PAIVA, E. S. DE., PEREIRA F. S. Capítulo 2 Deep Learning para Processamento de Linguagem Natural. Disponivel em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/86/378/642-1?inline=1>. Acesso em: 23 jan 2024.

PASCANU, R.; MIKOLOV, T.; BENGIO, Y. On the difficulty of training Recurrent Neural Networks. arXiv:1211.5063 [cs], 15 fev. 2013. DOI: https://doi.org/10.48550/arXiv.1211.5063

PENNINGTON, J., Socher, R., and Manning, C. D. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543, 2014. DOI: https://doi.org/10.3115/v1/D14-1162.

PURNELL, P. J. A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case study of SDG 13—Climate Action. Quantitative Science Studies, v. 3, n. 4, p. 1–27, 2022. DOI: https://doi.org/10.1162/qss_a_00215

RAHMAN, W. Inteligência Artificial e Aprendizado de Máquina. [S.l.]: Editora Senac São Paulo, 2022. ISBN 978-85-396-3278-7

RAMIRO, R. The Brazilian National Landscape in Governing and Implementing the 2030 Agenda and the SDGs. In: Galvao, T.G., de Menezes, H.Z. (eds) The Quest for the Sustainable Development Goals. Sustainable Development Goals Series. Springer, Cham, p. 3–18, 23 ago. 2024. DOI: https://doi.org/10.1007/978-3-031-59279-9_1

RAWORTH K (2017). Economia Donut: Uma alternativa ao crescimento a qualquer custo. Zahar. 363p.

RIBEIRO, H. F. Alinhamento das pesquisas de pós-graduação da Universidade Federal de Uberlândia com os Objetivos de Desenvolvimento Sustentável: Uma avaliação mediada por Inteligência Artificial. 2024. 96 f. Dissertação (Mestrado em Administração) - Universidade Federal de Uberlândia, Uberlândia, 2024. DOI http://doi.org/10.14393/ufu.di.2024.98

RIBEIRO, H. F. *et al*. ODS nas Universidades: Revisão sistemática utilizando Inteligência Artificial. In: SEMINÁRIO DE ADMINISTRAÇÃO – SEMEAD, 25., 2022, São Paulo. Anais [...]. São Paulo: FEA-USP, 2022. Disponível em: https://submissao.semead.com.br/25semead/anais/arquivos/1025.pdf. Acesso em: 6 jan. 2025.

ROMA, J. C. Os objetivos de desenvolvimento do milênio e sua transição para os objetivos de desenvolvimento sustentável. Ciência e cultura, v. 71, n. 1, p. 33–39, 2019. DOI: http://dx.doi.org/10.21800/2317-66602019000100011

ROMEIRO, A. R. Desenvolvimento sustentável: uma perspectiva econômico-ecológica. Estudos Avançados, v. 26, n. 74, p. 65–92, 2012. DOI: https://doi.org/10.1590/S0103-40142012000100006

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408, 1958. Disponível em: <https://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf>. Acesso em: 24 ago 2024.

RUSSELL, S.; NORVIG, P. Inteligência Artificial - Uma Abordagem Moderna. 4. ed. [s.l.]: Grupo GEN, 2022.

SACHS, J. D. *et al*. Six Transformations to achieve the Sustainable Development Goals. Nature Sustainability, v. 2, n. 9, p. 805–814, 2019. DOI: https://www.nature.com/articles/s41893-019-0352-9

SALVIA, A. L. *et al*., Assessing research trends related to Sustainable Development Goals: local and global issues. Journal of Cleaner Production, v. 208, p. 841–849, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0959652618329810>. Acesso em: 21 set. 2023.

SANTOS, A. B. dos. A terra dá, a terra quer. São Paulo: Ubu Editora/ PISEAGRAMA, 2023.

SANTOS, M.C.R. Desenvolvimento: construção e evolução do discurso político e teórico. In: SANTOS MCR. Guiné-Bissau: da independência colonial à dependência da Cooperação Internacional para o Desenvolvimento. Rio de Janeiro: Gramma. p. 25-46, 2021.

SANTOS, M. S. A formação regional do Extremo Sul da Bahia (1948–1974): desenvolvimento econômico, sociedade e meio ambiente. 2022. 395 f. Tese (Doutorado em Programa de Pós-Graduação em Estado e Sociedade) – Universidade Federal do Sul da Bahia, Porto Seguro, 2022.

SANTOS, Ê.; MORAES, L. E. ODSBAHIA-PTBR: A Natural Language Processing Model to Support Sustainable Development Goals. Revista de Gestão Social e Ambiental, São Paulo (SP), v. 18, n. 12, p. e010230, 2024. DOI: https://doi.org/10.24857/rgsa.v18n12-039.

SAVEGNAGO, C. L. *et al*., A AGENDA 2030 NAS UNIVERSIDADES FEDERAIS BRASILEIRAS: UM ESTUDO EXPLORATÓRIO, Humanidades & Inovação, v. 9, n. 14, p. 226–238, 2022.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. Neural Networks, v. 61, n. 61, p. 85–117, 2015. DOI: https://doi.org/10.1016/j.neunet.2014.09.003

SDG NEWS. Rainier Mallol: Revolutionizing public health with AI to predict disease outbreaks globally. 2024. Disponível em: <https://sdgnews.com/rainier-mallol-revolutionizing-public-health-with-ai-to-predict-disease-outbreaks-globally/>. Acesso em: 6 set. 2024.

SERAFINI, P. G. *et al*., Sustainable Development Goals in Higher Education Institutions: A systematic literature review. Journal of Cleaner Production, v. 370, p. 133473, 2022. Disponível em: <https://doi.org/10.1016/j.jclepro.2022.133473>. Acesso em: 24 jun. 2023.

SHARIFZADEH, M. *et al*. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. Renewable and Sustainable Energy Reviews, v. 108, p. 513–538, 2019. DOI: https://doi.org/10.1016/J.RSER.2019.03.040

SHAYAN, F.N *et al*. Sustainable Development Goals (SDGs) as a Framework for Corporate Social Responsibility (CSR). *Sustainability*. 2022; 14(3):1222. DOI: https://doi.org/10.3390/su14031222

SILVA, G. S. da ; ALMEIDA, L. A. Indicadores de Sustentabilidade para Instituições de Ensino Superior: Uma Proposta Baseada na Revisão de Literatura, Revista de Gestão Ambiental e Sustentabilidade, v. 8, n. 1, p. 123–144, 2019.

SOUZA, F., NOGUEIRA, R. & LOTUFO, R. (2020). BERTimbau: Pretrained BERT models for Brazilian Portuguese. Em: Intelligent Systems. Cham: Springer International Publishing, p. 403–417.

SRIVASTAVA, A. *et al*. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. 2022. DOI: https://doi.org/10.48550/arXiv.2206.04615

STERNBERG, R. J. Psicologia Cognitiva: tradução da 5ª edição americana. São Paulo: Cengage Learning, 2014.

TORRES, D. I. R. Contribución de la educación superior a los Objetivos de Desarrollo Sostenible desde la docencia. Revista Española de Educación Comparada, n. 37, p. 89, 2020. Disponível em: <https://doi.org/10.5944/reec.37.2021.27763>. Acesso em: 24 jun. 2023.

UNESCO. Artificial intelligence for sustainable development: challenges and opportunities for UNESCO's science and engineering programmes, 2019. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000368028> Acesso em 20 ago 2024.

UNESCO. Relatório de Ciências da UNESCO: A Corrida Contra o Tempo para um Desenvolvimento Mais Inteligente. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000377250_por. Acesso em: 20 de jan. 2025

UFSB - A UFSB: a universidade federal da Mata Atlântica. Ilhéus: UFSB, 2019. Disponível em: <https://ufsb.edu.br/component/content/article/1892-ufsb-a-universidade-federal-da-mata-atlantica>. Acesso em: 7 mar. 2025.

VASWANI, A. *et al*. (2017) Attention is all you need. In NIPS. DOI: doi.org/10.48550/arXiv.1706.03762.

VERGARA, S. C. Projetos e relatórios de pesquisa em administração. 5. ed. São Paulo: Atlas, 2004.

VINUESA, R. *et al*. The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, v. 11, n. 1, p. 233, 2020. DOI: https://doi.org/10.1038/s41467-019-14108-y

WANG, P. On Defining Artificial Intelligence. Journal of Artificial General Intelligence, v. 10, n. 2, p. 1–37, 2019. Disponível em: <https://sciendo.com/downloadpdf/journals/jagi/10/2/article-p1.pdf>. Acesso em: 7 jan. 2024.

WILMOTT, P. Machine Learning: An Applied Mathematics Introduction. [S.l.]: Panda Ohana Publishing, 2019. ISBN 978-1-9160816-0-4.

YE, Z. *et al*. Tackling environmental challenges in pollution controls using artificial intelligence: A review. The Science of the total environment, v. 699, n. 134279, p. 134279, 2020. DOI: https://doi.org/10.1016/j.scitotenv.2019.134279

ZHANG, C.; LU, Y. Study on Artificial Intelligence: The State of the Art and Future Prospects. Journal of Industrial Information Integration, v. 23, n. 23, p. 100224, 2021. DOI: https://doi.org/10.1016/j.jii.2021.100224

ZULAUF W. E. O meio ambiente e o futuro. Ambiente e desenvolvimento. Estud. av. 14(39). 2000. https://doi.org/10.1590/S0103-40142000000200009