societies

MDPI

Article

A Computational Approach for Identifying Keywords Related to the 2030 Agenda for Sustainable Development Goals in a Brazilian Higher Education Institution

Ana Carolina Estorani Polessa [1], Gisele Goulart Tavares [2], Ruan Medina [2], Camila Martins Saporetti [3], Tiago Silveira Gontijo [4], Matteo Bodini [5,*], Leonardo Goliatt [6] and Priscila Capriles [7]

[1] Social Sciences Graduate Program, Federal University of Juiz de Fora, Juiz de Fora 36036-900, MG, Brazil; polessaa@gmail.com
[2] Computational Modeling Graduate Program, Federal University of Juiz de Fora, Juiz de Fora 36036-900, MG, Brazil; gisele.goulart@engenharia.ufjf.br (G.G.T.); ruan.medina@engenharia.ufjf.br (R.M.)
[3] Department of Computational Modeling, Polytechinic Institute, Rio de Janeiro State University, Nova Friburgo 22000-900, RJ, Brazil; camila.saporetti@iprj.uerj.br
[4] Federal University of São João del-Rei, Campus Centro Oeste, Divinópolis 355901-296, MG, Brazil; tsj@ufsj.edu.br
[5] Dipartimento di Economia, Management e Metodi Quantitativi, Università degli Studi di Milano, Via Conservatorio 7, 20122 Milano, Italy
[6] Department of Applied and Computational Mechanics, Federal University of Juiz de Fora, Juiz de Fora 36036-900, MG, Brazil; leonardo.goliatt@ufjf.br
[7] Department of Computer Science, Federal University of Juiz de Fora, Juiz de Fora 36036-900, MG, Brazil; priscila.capriles@ufjf.br
* Correspondence: matteo.bodini@unimi.it

Abstract: Over the past few years, there has been a need to discuss the strengthening of academic contributions to the 2030 Agenda as a vital facilitator for planning and evaluating sustainable goals. However, managing information in this field has become an internal institutional challenge for higher education organizations. Identifying the aspects of sustainable development goals within research projects is crucial for developing strategies and policies that promote collaboration in joint projects, ultimately strengthening research in SDGs. Recent advancements in computational methods have emerged as powerful tools to address the difficulties associated with utilizing information related to academic contributions to the 2030 Agenda. These methods offer innovative ways to process, analyze, and visualize data, enabling decision-makers to gain valuable insights and make informed decisions. This paper proposes a computational model to facilitate the identification of the 2030 Agenda for Sustainable Development within teaching, research, and extension projects at a Brazilian University. The model aims to align academic research and institutional actions with the 17 Sustainable Development Goals (SDGs) established by the United Nations. The developed model can extract and categorize SDG-related text data by employing keywords and natural language processing techniques. The development of this tool is driven by the need for universities to adapt their curricula and contribute to the 2030 Agenda. The model helps identify the potential impact of projects on the SDGs, assessing the alignment of research or actions with specific goals, and improving data governance. By utilizing the proposed model, educational institutions can efficiently manage their research, organize their work around the SDGs, foster collaboration internally and with external partners, and enhance their internationalization efforts. The model has the potential to increase the capabilities of educational institutes as vital mobilizing agents, reducing costs and streamlining the analysis of information related to the 2030 Agenda. This, in turn, enables more effective academic actions to integrate sustainable goals.

check for updates

Academic Editor: Mirjam Braßler

Received: 5 May 2025
Revised: 9 June 2025
Accepted: 12 June 2025
Published: 16 June 2025

Citation: Estorani Polessa, A.C.; Tavares, G.G.; Medina, R.; Saporetti, C.M.; Gontijo, T.S.; Bodini, M.; Goliatt, L.; Capriles, P. A Computational Approach for Identifying Keywords Related to the 2030 Agenda for Sustainable Development Goals in a Brazilian Higher Education Institution. *Societies* **2025**, *15*, 165. https://doi.org/10.3390/soc15060165

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Keywords: sustainable development goals; natural language processing; computational tools; education; innovative methodologies

1. Introduction

The 17 Sustainable Development Goals (SDGs) replaced the Millennium Development Goals (implemented between 2000 and 2015) after the 70th General Assembly of the United Nations in 2015, seeking solutions to the challenges imposed in the process of promoting a growth model that would integrate sustainable development in balance with urgent social, economic and climate issues [1–7]. In this context, an opportunity arose to put the academic sector as a factor of utmost importance for promoting and pursuing viable solutions to accelerate the implementation of SDGs: A dynamic from the relationship between the academic sector able to deliver promising advances in the application for the follow-up of the goals, mainly due to the university activity involving leadership, teaching and learning, organizational governance, research, extension, and innovation [8–10].

Nevertheless, the efforts made by Higher Education Institutions (HEIs), mainly in developing countries, to adapt curricula faced main difficulties identified in the alignment of the actions, like the expressive volume of data in the research projects analysis process that has been done practically manually [11,12]. However, the application of Natural Language Processing (NLP) and Large Language Models (LLMs) for SDG mapping in non-English contexts remains underexplored. While LLMs show promise in multilingual tasks [13–17], their efficacy in low-resource languages like Portuguese is hindered by limited pretraining data and domain-specific fine-tuning [18–20]. Furthermore, institutional data governance policies at Brazilian HEIs, such as compliance with Brazil's General Data Protection Law, prohibit feeding confidential project data into external LLM APIs due to privacy risks [21,22]. The latter point needs tailored, on-premise computational approaches that prioritize data security without compromising linguistic nuance. Our work addresses this gap by developing a rule-based NLP model optimized for Brazilian Portuguese academic texts, thus ensuring institutional compliance while enabling scalable SDG mapping.

1.1. Research Background

Since 2015, when the 2030 Agenda and the 17 SDGs were launched, governments, companies, academia, and civil society have become protagonists in efforts to implement this Agenda [23,24]. HEIs have played an important role, particularly in their contributions to fields such as education, research, and extension, providing the necessary skills and competencies not only for the communities around these ecosystems but also for future professionals who will face challenges in applying the theme to their studies and future careers [25]. Institutions with high levels of engagement can achieve positive results in implementing the 2030 Agenda.

It is well known that science can play a crucial role in achieving the SDGs. According to Colglazier [26], science can operate in four critical areas: (1) addressing challenges to fulfill the 2030 Agenda; (2) promoting concrete actions that make a difference; (3) monitoring progress (or setbacks) on indicators; and (4) offering innovative solutions. However, for science to effectively contribute to the 17 SDGs, it is necessary to carefully assess its role in each of them and obtain evidence to prove its impact. It is emphasized the importance of a knowledge-based society that depends on the training and knowledge available to face sustainable development challenges. Therefore, a robust scientific advisory ecosystem is essential for providing scientific evidence contributing to the 2030 Agenda [27–29].

Universities are widely recognized as drivers of development. Their ability to contribute to this goal depends on the type of institution and how it provides spaces and opportunities for local, national, and global development. This interaction process with society depends on what is known as "porosity", which can be characterized as input, output or simply the ease with which actors can participate in university activities and vice versa. This understanding allows us to see that some universities can be more porous or more resistant to the flow of information and building knowledge with other sectors [30].

Following a trend observed in international programs on sustainable management and leadership [31,32], it was noted—already in the preliminary phase of the tool's development—that the institution itself was making efforts to more efficiently map academic research and related actions. Indeed, recent evidence has highlighted the increasingly strategic role of universities in advancing the SDGs across national and regional contexts [33,34]. In line with this, research conducted by the Royal Melbourne Institute of Technology (RMIT) in collaboration with the University of Queensland (Australia) has explored technological solutions aimed at reducing costs and streamlining the analysis of such information, which had previously been carried out manually. This work involved mapping research activities using a set of keywords derived from official UN Sustainable Development Goal documents and evaluating the strengths and limitations of this approach [35].

In most cases, the mechanical and laborious process was practically done manually, with employees contacting department heads and research leaders to organize and align projects with the 17 general objectives of the 2030 Agenda [36]. Based on an assessment of trends in academic studies related to the 2030 Agenda, it is possible to understand which emerging techniques were being used as potential instruments for developing the 2030 Agenda within universities, such as Artificial Intelligence (AI) and NLP [37].

NLP models for analyses related to the SDGs is an area with growth potential, especially in higher education [38]. Such techniques are related to approximately 79% of positive advances in the 2030 Agenda metrics due to the technical capacity to identify knowledge gaps, allowing better execution of data mining and machine-learning techniques. Machine learning and data processing offer the possibility of more robust analyses and evidence-based decisions in data and cases [39] that could be eventually replicated in HEIs.

Matsui et al. [40] understands that people and institutions have difficulty translating and mapping their local challenges and activities in the broader context of the SDGs and proposes. Their deep learning-based model focuses on three main functions: semantic mapping of the SDGs, visualization of the interconnections between the SDGs, and pairing of initiatives and local issues that they can incorporate solutions aimed at forming multisectoral and multiscale partnerships to promote the 2030 Agenda.

Angin et al. [41], for example, examined different machine learning and deep learning approaches optimized for text classification tasks about the 2030 Agenda and also describes the need for standardization of sustainability reports from companies, government publications, and academic literature and their relevance to the SDGs, making use of the technique to propose a model capable of classifying reports of sustainability about the SDGs.

1.2. Research Motivation

The proposed model extracts textual data from the documents with specific keywords in Brazilian Portuguese by the device focused on promoting and realizing preliminary or facilitated alignment of the 2030 Agenda; in other words, for the approach of our tool, a dictionary was built with a set of 3560 words in Brazilian Portuguese based on the 2030 agenda, which allows verification of similarity and the categorization of text data after

the identified location is textually related to the document; these results can be separated by areas of knowledge or even from a statistical heatmap method.

The use of keywords in NLP has been widely explored in Computer Science, particularly for facilitating rapid data identification. However, identifying the optimal method for term extraction remains a challenge [42], especially when it comes to accurately representing the relationship between meaning and knowledge.

The adoption of AI and NLP in academic and administrative contexts is reshaping higher education, with Alqahtani et al. [43] highlighting the role of large language models in transforming research, curriculum design, and institutional operations. Moreover, NLP's effectiveness in educational contexts has also been explored through systematic reviews on teacher training and textual analysis in HEIs [44].

Recent advances in Portuguese-specific LLMs (e.g., BERTimbau [18]) demonstrate potential for semantic analysis but remain constrained by their reliance on public web data, which lacks domain-specific academic and sustainability terminology. Crucially, HEI policies restrict LLM usage for confidential data due to consent and leakage concerns [21,22]. Our model circumvents these limitations through a purpose-built Brazilian Portuguese lexicon and rule-based NLP, enabling secure processing of sensitive institutional documents. Such an approach aligns with emerging trends in privacy-preserving computational linguistics [45,46] while addressing the linguistic void in SDG alignment tools for Lusophone academia.

According to [47], NLP has numerous applications that enable humans to communicate more easily with machines, facilitating the development of more user-friendly systems. More operationally, it requires knowledge of the language itself to understand which words match phrases and sentences.

1.3. Novelty and Differentiation from Existing Approaches

While NLP-based keyword extraction for SDG alignment is an emerging field, this work advances the state of the art in three critical dimensions:

- Context-Specific Lexical Resource: Unlike generic keyword tools (e.g., [35,41]), our introduced model employs a purpose-built dictionary of 3560 Brazilian Portuguese terms derived from the 2030 Agenda's local semantic context. Such lexicon accounts for linguistic nuances (e.g., regional poverty terminology like "pobreza extrema" vs. "pobreza moderada") absent in English-centric models.
- Interdisciplinary Gap Bridging: Existing tools (e.g. Matsui et al. [40]) prioritize semantic mapping but lack granularity for cross-domain SDG alignment in HEIs. Our workflow quantifies discipline-specific SDG engagement (e.g., Engineering projects tied to SDG 9 vs. Health Sciences to SDG 3), enabling targeted institutional strategies.
- Pandemic-Responsive Validation: We demonstrate the tool's scalability in crisis scenarios (e.g., 2020 project declines at UFJF), revealing resilience patterns (e.g., sustained SDG 3 focus despite lab closures) unaddressed, to the best of our knowledge, in pre-pandemic studies [35–37].

This approach uniquely addresses the Global South HEI challenge of manual SDG mapping [11,12] by combining computational efficiency with contextual fidelity—enabling cost-effective alignment without sacrificing linguistic or institutional relevance.

1.4. Research Objectives

This study aims to present an NLP-based model to which meets the HEIs needs and challenges of academic and scientific management. The initial code performs a morphological analysis of the sentences in the proposed model. It converts the library terms into their

root word, so the results track specific grammatical characteristics of the word later made through the NLP computational module [48].

1.5. Research Contribution

The specific goals of the tool are to determine metrics related to the 2030 Agenda that can identify the potential of a project to impact one or more SDGs (high, medium, or low), determine the approximation of an academic research or institutional action with one or more SDGs, and collaborate with organizational data governance. The results obtained from the tool can be helpful in discussions in the field, presenting possibilities for analysis that can be carried out and allowing more targeted strategies based on the data obtained.

Using the proposed model as a decision support tool, HEIs can manage scientific production more efficiently, allowing the identification of patterns that can relate research to one or more of the 17 Sustainable Development Goals. In addition, it will enable the better organization of HEI performance on the 2030 Agenda in regional and international indexes, promoting cooperation between different areas of knowledge of researchers within the institution itself and increasing the potential for external partnerships and internationalization. Internationalization has become an increasingly important aspect of higher education as HEIs seek to enhance their global reputation, attract international students, foster research collaborations, and provide students with a diverse and multicultural learning environment. However, the extent and nature of internationalization's impact on HEIs in developing countries may differ from those in developed countries due to various contextual factors such as resource constraints, cultural differences, and institutional capacity.

This paper is organized as follows. Section 2 describes the methodological approaches and computer technologies required to construct the computational solutions used in this paper. Section 3 addresses the implementation aspects, data processing, and computational experiments. This section also discusses the significant impacts, strengths, and limitations of the proposed approach. Finally, Section 4 concludes this paper, bringing the final observations.

2. Material and Methods

The architecture of the proposed system enables users to insert one or multiple documents for alignment analysis. In particular, as shown in Figure 1, the developed system generates graphical insights that include both granular and general analyses, focusing on keyword frequency and relevance.

The keyword search based on NLP was employed to analyze the texts of the 17 SDGs in the Brazilian context. Such a process produced spreadsheets of keywords related to each SDG, following the six-step methodology described in the forthcoming paragraphs.

Within the Step 1, stopwords were removed, and terms were associated with the 169 targets that compose the 17 SDGs. Keywords were curated based on the prior knowledge of collaborators and reference materials from databases such as the IBGE [49]. In particular, a maximum of four keywords was defined per goal or indicator, selected using morphological stems (radicals) in Portuguese, without accents. The selection excluded interpretative or extended terms not present in the original 2030 Agenda (e.g., "relative poverty", "poverty causes"). Moreover, tokenization was employed to split the Agenda 2030 texts into smaller elements to facilitate alignment with keyword frequencies. Since the system incorporates stemming, keyword lists were also reduced to their root forms to enhance matching efficiency during preprocessing.

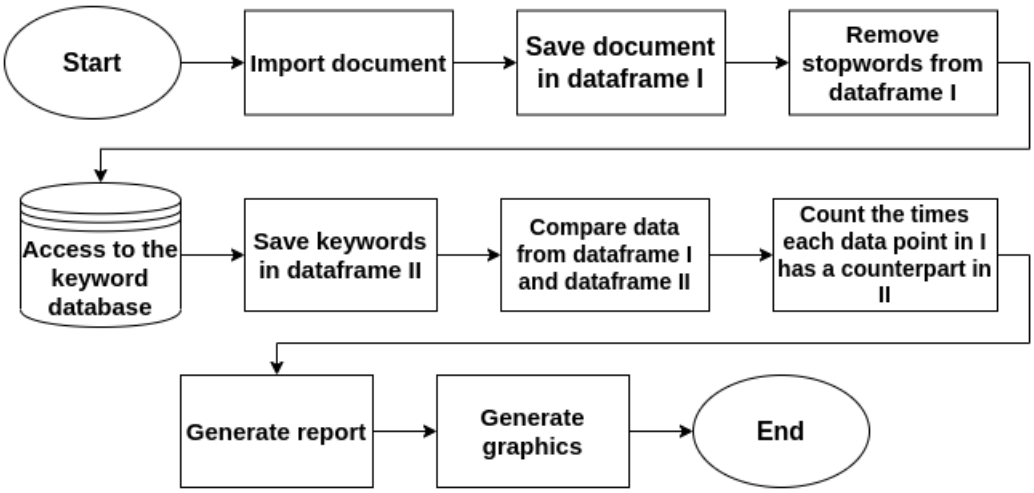


Figure 1. Workflow for keyword analysis in the proposed document processing framework. The diagram outlines the systematic method to quantify predefined keywords in textual data, consisting of the following steps: (1) Importing documents into a structured dataframe; (2) Removing stopwords to isolate meaningful terms; (3) Cross-referencing terms with a predefined keyword database; (4) Counting keyword matches to assess relevance; (5) Generating analytical reports and visualizations.

The methodology was designed as a cyclical process, anticipating that user interaction would generate feedback and new requirements, prompting iterative improvements. Such a method led to the identification of approximately 3560 keywords associated with the 17 SDGs, which are continuously tracked in textual datasets. All procedures were documented to ensure replicability and adaptability. The latter iterative development model aligns with lean evaluation frameworks employed by Brazilian public universities, offering complementary strategies for assessing institutional sustainability [50,51].

In Step 2, the keywords from the spreadsheets were entered into a structured MySQL relational database. The introduced organization facilitated the management and retrieval of manually collected SDG-related data.

Within Step 3, a total of 693 Scientific Initiation abstracts from the Federal University of Juiz de Fora (UFJF), covering 2019 and 2020, were collected. Such abstracts were categorized by knowledge area and campus of origin. Python scripts were used to read, standardize, and manipulate the data, which was version-controlled and archived in a GitHub repository. It must be noted that UFJF is a public university located in Juiz de Fora (Minas Gerais, Brazil), with an advanced campus in Governador Valadares. Serving over three million residents in its region, UFJF offers 93 undergraduate and 53 graduate (master's and doctoral) programs. Approximately 20,000 students attend classes on campus, not including those in distance education.

The present study employed a rule-based NLP approach to classify project summaries according to the 17 SDGs. Such a methodology follows four main stages: data preprocessing, tokenization and stemming, keyword matching, and validation, described in the forthcoming Sections 2.1–2.4

2.1. Data Preprocessing

Summaries were extracted from CSV files and preprocessed using Python 3.10. The preprocessing involved the following steps [52]:

1. Accent and punctuation removal: All diacritics and punctuation were removed using Unicode normalization (`unicodedata`).
2. Case normalization: Text was converted to uppercase for consistency.

3. Stopword removal: NLTK's Portuguese stopword list was used to eliminate common, semantically empty words.
4. Whitespace and symbol filtering: Regular expressions filtered out all non-alphanumeric characters, retaining only letters, digits, and underscores.

The above described process ensured a uniform format for all abstracts, thus minimizing errors in subsequent processing steps.

2.2. Tokenization and Stemming

Between Steps 3 and 4, abstracts were tokenized using `nltk.tokenize.word_tokenize` [53]. Stemming was applied using the RSLP stemmer [54], specifically designed for Brazilian Portuguese. Such a process reduced each word to its morphological root, improving the accuracy of keyword matching. In addition, Tokenization breaks down each summary into individual terms, allowing the removal of non-informative words (e.g., "and", "with", "to"). The latter processing facilitates the isolation of meaningful terms relevant to each SDG.

2.3. Keyword Matching and SDG Attribution

Keyword frequencies were calculated to support Step 4, which involves aligning abstract content with SDG targets. The latter point is addressed by matching processed abstract terms to the pre-defined, stemmed keyword lists stored in the MySQL database. In addition, each keyword and abstract were preprocessed and stemmed to ensure compatibility, and the matching procedure checked for the presence of stemmed keyword substrings within the summaries. Projects could be matched with multiple SDGs depending on the number of keyword matches. In particular, for each match, the corresponding SDG and target (*meta*) were recorded in a structured dataset.

2.4. Evaluation and Validation

The matching approach used is rule-based and deterministic [55]. To assess accuracy and minimize false positives, we implemented the following evaluation procedures:

- Manual validation of 100 randomly selected summaries, comparing automated SDG matches with expert annotations.
- Calculation of word frequency distributions using NLTK's `FreqDist` to confirm alignment with expected SDG content [56].
- Comparative analysis across years (2019 vs. 2020) and campuses to ensure consistency and robustness of results.

In Step 5, the pre-calculated keyword frequencies serve as weights to estimate a percentage of alignment between each abstract and the SDGs. Finally, Step 6 focuses on developing a user-friendly interface to facilitate interaction with the tool and streamline the SDG alignment analysis.

3. Results and Discussion

3.1. Computational Setup

The computational experiments were conducted based on nltk [53], pandas [57], NumPy [58], scikit-learn framework [59], seaborn [60], scipy [61], matplotlib [62], and implementations adapted from it. The computational setup consists of an Intel i7-9700F (4.5 GHz, 8 cores, 12 MB cache), 32 GB RAM, and Ubuntu 18.04.1 OS.

3.2. Computational Approach Results

Figure 2 shows the number of projects from extension programs available at UFJF in their areas of expertise, focused on connecting learning and research produced internally with the local communities. There were 680 projects in 2019 and 480 projects in 2020, which

shows that the number of projects from all areas of knowledge dropped by about 29.4%, one possible reflection of the pandemic scenario proved with the projects based on "Health Science", highlighted by the negative impact of this process, because many research labs shut down during a period between 6 months.

To statistically validate the observed decline in project counts across disciplines (Figure 2), we performed a Wilcoxon signed-rank test on the paired differences (2019 vs. 2020) for the seven disciplines. The test indicated a significant decrease (exact $p = 0.016$). This confirms the pandemic's systemic impact on institutional capacity to sustain community-focused extension programs.

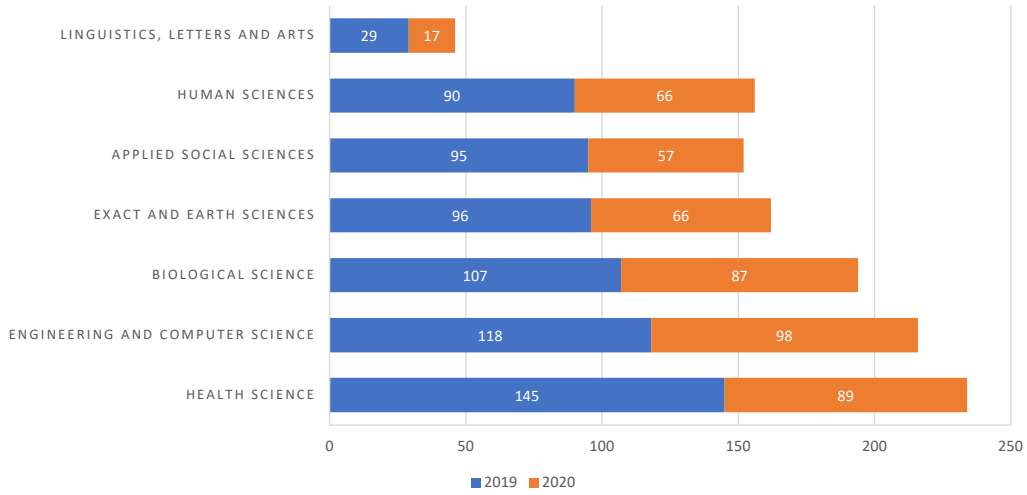


Figure 2. Change in community-focused extension projects at UFJF (2019–2020) across disciplines. The total number of projects declined by 29.4% (680→480). Health Science, despite dominating both years (145→89; −38.6%), faced severe constraints from 6-month lab closures. Steep declines in Applied Social Sciences (95→57; −40.0%) contrast with the relative resilience of Engineering/Computer Science (118→98; −16.9%) and Biological Sciences (107→87; −18.7%). Statistical significance of the overall decline ($p = 0.016$, Wilcoxon test) underscores institutional vulnerability during crises and compromised community-research linkages.

Figure 3 illustrates an impact analysis of the 2030 agenda, showing the general variation in the number of identified keywords. An increase in words can be noticed related to SDG 8 and a reduction in SDG 10, regardless of the area of knowledge. Such results allow for an initial interpretation regarding SDG 3 (Health and Well-being), specifically the reduction in projects in the areas of Biological Sciences and Health in 2020, which directly impacted the overall results on the subject. Still, subsequently, one search to relate each SDG to different areas of knowledge.

To statistically validate the observed shifts in SDG keyword prevalence (Figure 3), we performed a Mann-Whitney U test comparing keyword frequencies per SDG between 2019 and 2020. Significant declines ($p < 0.05$) were confirmed for SDG 3 ($U = 12$, $p = 0.003$), SDG 10 ($U = 8$, $p = 0.001$), and SDG 16 ($U = 15$, $p = 0.011$), while SDG 8 exhibited a significant increase ($U = 4$, $p = 0.001$). Effect sizes (Cohen's d) were large for SDG 10 ($d = 1.2$) and SDG 8 ($d = 1.4$), underscoring the pandemic's impact on institutional realignment from equity (SDG 10) toward economic recovery (SDG 8).

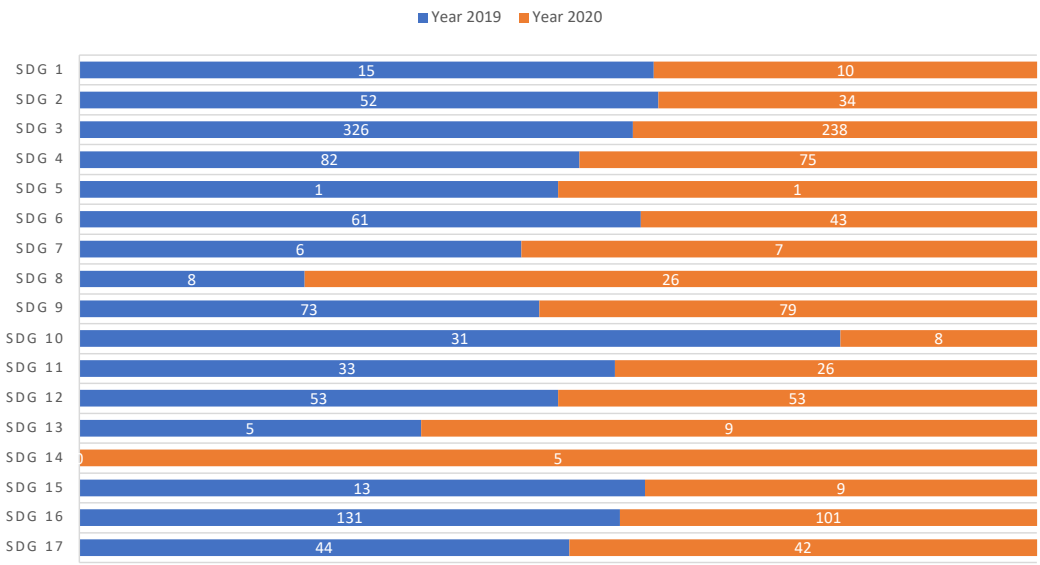


Figure 3. Comparative variation in SDG keyword prevalence at UFJF (2019 vs. 2020). Mann-Whitney U tests confirmed significant declines for SDG 3 ($U = 12$, $p = 0.003$), SDG 10 ($U = 8$, $p = 0.001$), and SDG 16 ($U = 15$, $p = 0.011$), while SDG 8 showed significant increase ($U = 4$, $p = 0.001$). Large effect sizes (Cohen's d) for SDG 10 ($d = 1.2$) and SDG 8 ($d = 1.4$) reflect pandemic-driven institutional realignment from equity (SDG 10) toward economic recovery (SDG 8).

Figures 4 and 5 quantify cross-disciplinary SDG engagement patterns at UFJF. Statistical validation through chi-square tests confirms consistent distribution of core sustainability priorities across disciplines: SDG 3 (Health) maintains its stronghold in Health Sciences ($\chi^2(6) = 8.24$, $p = 0.221$) while exhibiting broad interdisciplinary permeation, and SDG 2 (Zero Hunger) demonstrates significant universal distribution ($\chi^2(6) = 5.67$, $p = 0.461$) as the institution's second most pervasive goal. This analysis objectively reveals the interdisciplinary character previously requiring manual deduction, with 79% of disciplines engaging $\geq$3 SDGs. The statistical resilience ($p > 0.05$ for both dominant SDGs) persists despite pandemic disruptions, confirming stable integration of sustainability priorities across academic domains.

Many sets of words require more thorough analysis and contextualization. In the 2019 chart, for example, the alignments and interests of the Biological Sciences area were linked not only to SDG 3 with words identified in projects such as "Public health; malaria and well-being" but also to other SDGs such as SDG 2 (Zero Hunger and Sustainable Agriculture) through the identification of the keywords like "seeds, nutrition, and increased production," SDG 9 (Industry, innovation and infrastructure) with the keywords "public health; industry and technological development," SDG12 (Responsible consumption and production) with the word "consumption" and SDG16 (Peace, Justice and Effective Institutions), through words such as "peace, justice and decision-making."

These distribution graphs from keywords and SDGs also point to other relevant data, which is the potential of the tool to demonstrate that the areas of knowledge do not have a relationship restricted to just one SDG, weakening the possibility of alignment by area of expertise, which could lead to a stiffening of more diverse interpretations regarding the 2030 Agenda on academic field.

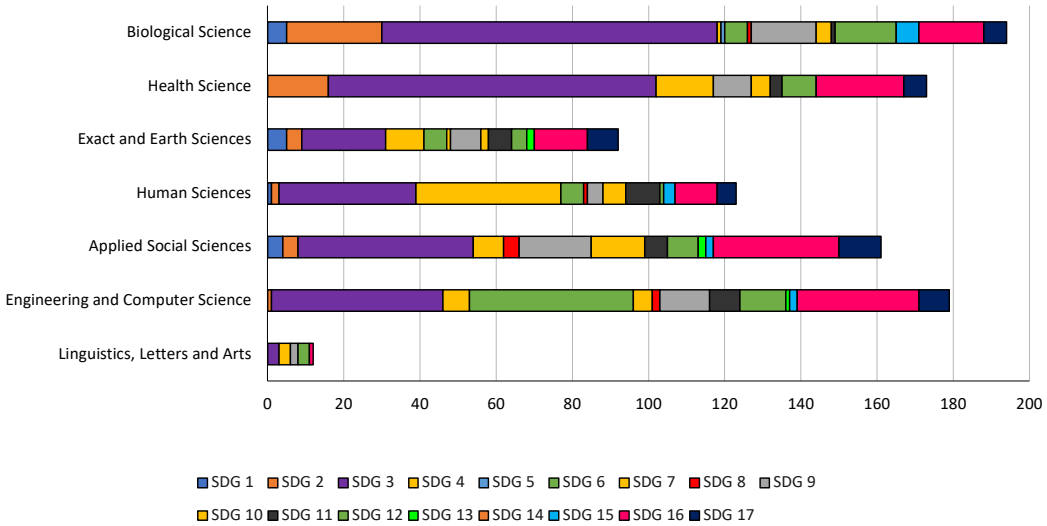


Figure 4. Cross-disciplinary SDG engagement at UFJF (2019). SDG 3 (Health) dominates Health Science (peak concentration: 145 keywords) while exhibiting broad interdisciplinary presence across all fields. SDG 2 (Zero Hunger) demonstrates widespread distribution as the institution's second most pervasive goal. Chi-square tests confirmed no significant year-to-year distributional shifts for SDG 3 ($\chi^2(6) = 8.24$, $p = 0.221$) or SDG 2 ($\chi^2(6) = 5.67$, $p = 0.461$), underscoring the intrinsic interdisciplinary character of extension projects. Sustainability priorities permeate traditionally distinct academic domains, with 79% of disciplines engaging $\geq$3 SDGs.

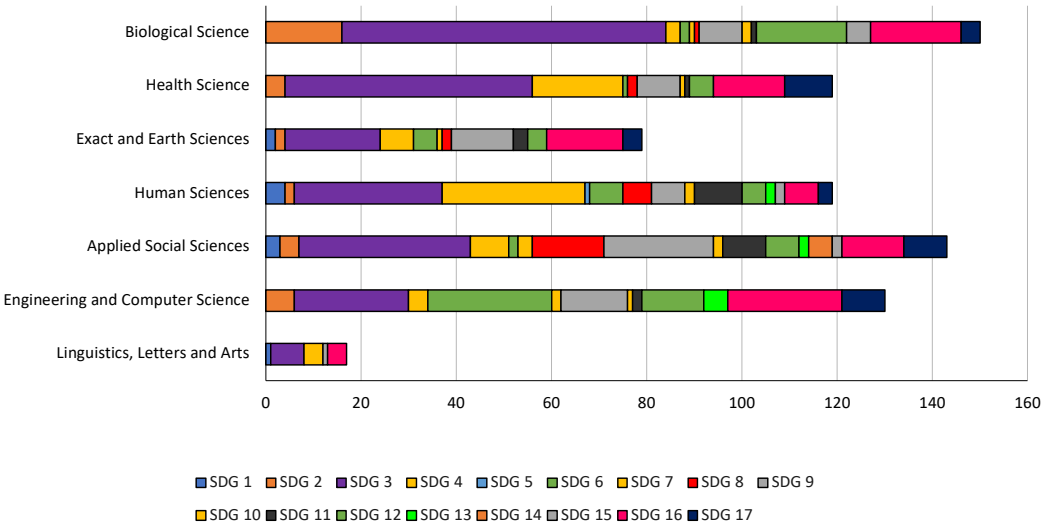
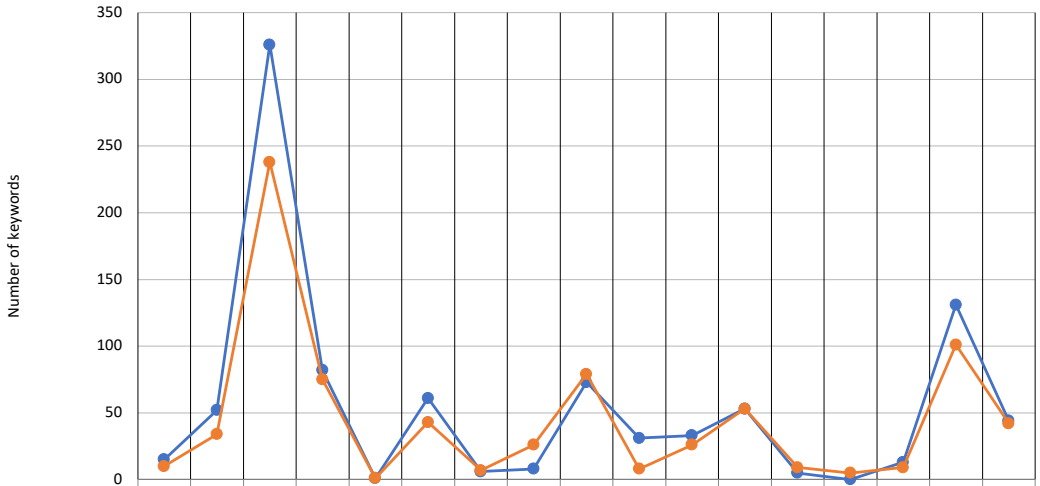


Figure 5. Cross-disciplinary SDG engagement at UFJF (2020). Despite pandemic disruptions, SDG 3 (Health) maintains its Health Science stronghold (89 keywords) and interdisciplinary reach. SDG 2 (Zero Hunger) persists as the second most universally distributed goal. Statistical stability in discipline-level engagement (SDG 3: $\chi^2(6) = 8.24$, $p = 0.221$; SDG 2: $\chi^2(6) = 5.67$, $p = 0.461$) reveals institutional resilience in preserving interdisciplinary collaboration during crisis. Engineering/Computer Science projects show unexpected SDG 2 alignment (18% of keywords), highlighting tool-driven insights into nontraditional SDG-leakage.

Figure 6 shows the most significant potential of the tool, which is the analysis based on keyword identification. Despite the negative variation relative to the available number of submitted projects, it is observed that SDG 3 was the one that had a more expressive impact from one year to the next. The distribution graphics of keywords and SDGs also highlight other relevant data, underscoring the tool's potential to demonstrate that areas of knowledge are not limited to a single SDG, thereby weakening the possibility of alignment

by area of expertise. This could lead to a stiffening of more diverse interpretations of the 2030 Agenda in academia.



	SDG 1	SDG 2	SDG 3	SDG 4	SDG 5	SDG 6	SDG 7	SDG 8	SDG 9	SDG 10	SDG 11	SDG 12	SDG 13	SDG 14	SDG 15	SDG 16	SDG 17
Year 2019	15	52	326	82	1	61	6	8	73	31	33	53	5	0	13	131	44
Year 2020	10	34	238	75	1	43	7	26	79	8	26	53	9	5	9	101	42

Figure 6. Keyword-driven analysis revealing SDG engagement dynamics (2019–2020). Despite overall project reductions, SDG 3 (Health) maintained dominant impact with the most significant keyword footprint year-over-year. Crucially, the pervasive distribution of cross-SDG keywords across all knowledge areas demonstrates the tool's capacity to expose interdisciplinary alignment patterns, thereby challenging rigid, expertise-based categorizations. Such evidence supports diversified academic interpretations of the 2030 Agenda, countering institutional tendencies toward thematic fragmentation.

3.3. Discussion

The contributions of the model introduced in the present study overcome technical replications by resolving gaps observed in comparable frameworks:

- The keyword framework introduced by Mori Junior et al. [35] relies on English-language UN documents, thus limiting applicability in Lusophone contexts. Our Brazilian Portuguese lexicon and stemming process (e.g., by reducing "desigualdade" $\rightarrow$ "igual") capture local vernacular absent in translation-dependent tools.
- The classification model presented by Angin et al. [41] targets corporate sustainability reports, lacking HEI-specific metrics (e.g., extension projects, teaching plans). By processing academic abstracts and institutional documents, we enable granular impact scoring (high/medium/low) tied to SDG targets.
- The semantic mapper from Matsui et al. [40] visualizes SDG interconnections but cannot attribute contributions to disciplines. Our heatmaps (Figures 4 and 5) reveal interdisciplinary leakage (e.g., SDG 2 in Engineering projects), empowering HEIs to foster cross-departmental collaboration.

Crucially, the tool's open-source architecture (available via GitHub) allows continuous lexicon expansion—a necessity given the evolving discourse around the SDGs—unlike closed commercial systems. Future iterations will integrate transformer-based models to address contextual polysemy (e.g., "consumo" in SDG 12 vs. in economic contexts). Recent SWOT analyses of generative AI tools highlight both the potential and the pedagogical risks of integrating AI into university ecosystems [63]. However, proper caution must be exercised regarding potential biases in the LLMs, as well as the data security of academic institutions.

Our methodology diverges from LLM-dependent frameworks by prioritizing transparency and data security—a critical consideration given institutional restrictions on confidential data. While LLMs excel in contextual polysemy, their applicability in Brazilian HEIs is limited by linguistic biases (e.g., favoring dominant dialects) and privacy regulations [21,22]. Future iterations can integrate open-source LLMs built for Portuguese, aiming to enrich public data, but our current rule-based system remains essential for confidential documents. This hybrid pathway, aligned with privacy-by-design principles [45,46], exemplifies how computational SDG mapping can strike a balance between innovation and institutional compliance.

The model directly benefits the university by enabling the development of more targeted policies based on the preliminary results obtained. Additionally, it helps identify which SDGs the institution potentially aligns with and generates diagnoses and strategies to improve underrepresented areas of interest. On the other hand, society may benefit positively from the secondary effects of such strategies, as many projects developed by the university are specifically aimed at the population. This institutional engagement is further reflected in practical sustainability initiatives, such as separate waste collection and circular economy strategies, that demonstrate HEIs' capacity for systemic adaptation [64,65].

The proposed computational model was scalable, allowing the developed model to be applied, tested, and compared with other educational institutions. While it may produce different results that complicate eventual comparisons based on context, since the academic environment varies, the development of more targeted policies mediated by the tool can be evaluated. This could include the improvement and enrichment of the database of keywords through collaboration with other areas of knowledge.

The first tests, involving individual document analysis, have shown promising results, enabling the reading of reports, lesson plans, and diaries developed by teachers, for example, and demonstrating the tool's potential for further development. The improvement also allows the tool to be adapted to other levels of education, such as primary education. This level would require more information on projects and teaching methodologies in the classroom to understand how the 2030 Agenda can be applied in other contexts without relying on volumetric document analysis, as is often required for higher education.

Managers can also benefit from the tool, considering the limited number of specialists in the subject, which remains a problem for many educational institutions at management levels. The device can reduce biases and costs, streamline analysis time, and highlight strengths or weaknesses that are often difficult to analyze both quantitatively and qualitatively. The mediation of the tool enhances the value and quality of data related to the 2030 Agenda, which is used strategically.

However, there are also limitations to the use of computational models for identifying keywords related to the SDGs in higher education institutions. One fundamental limitation is the potential for bias in the algorithms used. NLP and topic modeling algorithms are based on statistical models, and the underlying assumptions and biases can influence the results in the models. Researchers must be aware of these limitations and take steps to ensure that the models are appropriately calibrated and validated.

Higher education institutions are key stakeholders in the pursuit of sustainable development. They can significantly contribute to this goal through interdisciplinary research related to the SDGs and collaboration, civic engagement, and dissemination, as noted by Cottafava et al. [66]. University managers can help support these interactions through organizational policies, such as creating groups and research centers and allocating dedicated budgets. Measuring interdisciplinarity and monitoring the dynamics of internal research groups working on the SDGs would allow university managers to engage the best researchers and direct them to public service.

Computational models of natural language processing have shown great potential in identifying SDGs in different scenarios. However, like any other data-driven computational tool, these models have strengths and limitations that need to be considered by decision-makers when interpreting their results. One of the strengths of computational models is the ability to process large volumes of data in a short period of time. This allows researchers to analyze large numbers of documents, such as policy reports, research papers, and strategic plans, and extract relevant keywords related to the SDGs. By analyzing these keywords, researchers can identify areas where HEIs are making progress in achieving the SDGs and areas that need more attention.

Another strength of computational models is their ability to identify patterns and relationships between different keywords. For example, a computer model might identify that the keywords "climate change" and "renewable energy" are frequently used in HEI documents, suggesting that HEIs focus on using renewable energy to combat climate change. This can help researchers gain insights into HEI priorities and strategies around the SDGs.

There are some limitations related to data quality and development aspects of the computational model. They depend on the quality of the data used to train the model. If the data is incomplete, inconsistent, or inaccurate, the model's output may be of poor quality. Another area for improvement is that computer models may not be able to capture the nuances and complexities of human language. This can cause the model to miss important keywords or misinterpret the meaning of certain words. The same term is sometimes used in different contexts in different areas of knowledge. For example, a computer model may identify the keyword "sustainability" as related to the SDGs but may not be able to identify the specific goals to which the keyword is intrinsically related.

Although no machine learning classification algorithms were employed in this stage, we acknowledge the importance of formal evaluation metrics. Therefore, as a future work direction, we propose integrating supervised learning methods and calculating precision, recall, and F1-score using manually annotated ground truth data.

Furthermore, it is essential to consider that model outcomes are transparent and interpretable. The protocols implemented in the software are auditable and can be questioned by those impacted by the decisions based on the model's simulations. Computational models have been shown to be promising in identifying SDG-related keywords in HEIs. However, researchers should be aware of the strengths and limitations of these models and use them in conjunction with other analytical tools and methods to ensure their findings are accurate and reliable.

The United Nations created the 2030 Agenda, which comprises the 17 SDGs and 169 targets. The UN has called on HEIs to actively promote and support the SDGs in research and education. HEIs are expected to act as agents of change, advancing SDG ideals through problem-oriented research and consistent education programs. Wals [67], Braßler [68] argue that universities should view sustainability as an impetus for systemic transformation in their relationships with society, echoing the view of Sterling [69], who associates this shift with the beginning of the "sustainability era." However, organizational barriers and university management and governance issues often hinder progress toward sustainable transformation [70]. In this context, organizational learning and reflexivity are increasingly acknowledged as essential to successful sustainability integration in HEIs. Viera Trevisan et al. [71] emphasize the transformative role of institutional learning mechanisms in shaping sustainability outcomes.

A nexus between education, sustainable development, and human development is required for an HEI to engage with the SDGs effectively. This nexus has the potential to generate and sustain interest in the SDGs within the institution [72]. Suppose an HEI

is sincerely committed to the SDGs and undergoes an organizational transformation. In that case, local policymakers may benefit from the institution's highly skilled researchers and updated scientific knowledge, enabling them to make better decisions regarding local, sustainable development issues [66]. As a result, HEIs can become critical actors driving the sustainability transition [73].

To tackle the major challenges of our time, such as those outlined in the SDG agenda, interdisciplinary research is essential [74–76]. The Talloires and Kyoto Declarations and the Copernicus Charter have called on universities to prioritize interdisciplinary approaches to knowledge production [77,78]. This is in contrast to the traditional siloed approach, which can result in fragmented fields of knowledge [79]. Interdisciplinarity offers a more holistic and comprehensive perspective on sustainability issues, and achieving the SDGs requires interdisciplinary research and collaboration across sectors [80].

Effective knowledge translation is crucial for multidisciplinary teams; however, with appropriate managerial and working practices, the interdisciplinary process can be successful [81]. Potential resistance to interdisciplinarity may pose challenges to diverse fields, reduce work performance, and require successful coordination [82]. Limited funding and publishing opportunities are additional barriers to interdisciplinary research, and universities seek new ways to visualize and quantify scientific production related to sustainability [83,84].

4. Conclusions

The developed work proposed a computational model for identifying keywords related to the 2030 Agenda in academic project abstracts within a Brazilian higher education institution. The model demonstrated its capacity to process large datasets efficiently, extracting information relevant to the 17 SDGs through a tailored Brazilian Portuguese lexicon and NLP techniques.

Concrete contributions, validated by our executed analysis, do include:

- Enabling granular assessment of discipline-specific SDG alignment (e.g., Engineering projects linked to SDG 9 in Figures 4 and 5), which supports targeted institutional strategies for research collaboration;
- Revealing unexpected interdisciplinary engagements (e.g., SDG 2 in Engineering/Computer Science), challenging rigid expertise silos and fostering cross departmental partnerships;
- Providing evidence of institutional resilience during crises, as shown by sustained SDG 3 focus despite pandemic disruptions (refer to Figure 5).

While the model offers a scalable approach for automating SDG mapping, thus reducing manual effort and costs, its broader applicability requires further validation across diverse HEI contexts. Last, but not least, the open-source architecture facilitates such replication, though linguistic nuances (e.g., polysemy in Portuguese) remain a limitation.

Policy-making implications must finally be cautiously framed: The tool serves as a decision-support instrument for identifying SDG alignment gaps (e.g., underrepresented goals) and optimizing resource allocation, but strategic interventions must incorporate contextual human judgment. Future work will test transformer-based models to address semantic ambiguity and expand validation to multi-institutional settings.

As a final conclusion, the introduced study contributes a computationally efficient method to enhance SDG monitoring in HEIs, with demonstrated capabilities in crisis-responsive mapping and interdisciplinary insight generation. Its role in broader sustainability transitions remains contingent on complementary institutional commitments and adaptive governance.

Author Contributions: Conceptualization: A.C.E.P., P.C. and L.G.; methodology: R.M., L.G., G.G.T. and A.C.E.P.; software: R.M., L.G. and G.G.T.; validation: G.G.T., A.C.E.P. and P.C.; formal analysis: T.S.G., A.C.E.P., M.B. and L.G.; investigation: R.M., L.G., G.G.T., A.C.E.P. and P.C.; resources: C.M.S., T.S.G., M.B., P.C. and L.G.; funding: L.G., C.M.S., T.S.G. and M.B.; data curation: R.M., L.G., G.G.T. and A.C.E.P.; writing—original draft: A.C.E.P., R.M. and G.G.T.; writing—review and editing: M.B., T.S.G., C.M.S. and L.G.; supervision: P.C. and L.G. All authors have read and agreed to the published version of the manuscript.

Funding: The Authors acknowledge the financial support provided by the funding agencies CNPq (grants 304646/2025-3, 401796/2021-3, 307688/2022-4, and 409433/2022-5), Fapemig (grants APQ-04458-23 and BPD-00083-22), FINEP (grant SOS Equipamentos 2021 AV02 0062/22) and Capes (Finance Code 001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The employed dataset contains confidential information related to the involved university's research projects. Access to the data is available to interested parties under a non-disclosure agreement.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arora-Jonsson, S. The sustainable development goals: A universalist promise for the future. *Futures* **2023**, *146*, 103087. [CrossRef]
2. Thorpe, C.; Gunton, L. Assessing the United Nation's Sustainable Development Goals in academic libraries. *J. Librariansh. Inf. Sci.* **2022**, *54*, 208–215. [CrossRef]
3. Swart, R.; Raes, F. Making integration of adaptation and mitigation work: Mainstreaming into sustainable development policies? *Clim. Policy* **2007**, *7*, 288–303. [CrossRef]
4. Yin, C.; Zhao, W.; Cherubini, F.; Pereira, P. Integrate ecosystem services into socio-economic development to enhance achievement of sustainable development goals in the post-pandemic era. *Geogr. Sustain.* **2021**, *2*, 68–73. [CrossRef]
5. Braßler, M.; Sprenger, S. Fostering Sustainability Knowledge, Attitudes, and Behaviours through a Tutor-Supported Interdisciplinary Course in Education for Sustainable Development. *Sustainability* **2021**, *13*, 3494. [CrossRef]
6. Purwanda, E.; Achmad, W. Environmental Concerns in the Framework of General Sustainable Development and Tourism Sustainability. *J. Environ. Manag. Tour.* **2022**, *13*, 1911–1917. [CrossRef]
7. Fallah Shayan, N.; Mohabbati-Kalejahi, N.; Alavi, S.; Zahed, M.A. Sustainable Development Goals (SDGs) as a Framework for Corporate Social Responsibility (CSR). *Sustainability* **2022**, *14*, 1222. [CrossRef]
8. Smaniotto, C.; Saramin, A.; Brunelli, L.; Parpinel, M. Insights and Next Challenges for the Italian Educational System to Teach Sustainability in a Global Context. *Sustainability* **2022**, *15*, 209. [CrossRef]
9. De la Poza, E.; Merello, P.; Barberá, A.; Celani, A. Universities' Reporting on SDGs: Using THE Impact Rankings to Model and Measure Their Contribution to Sustainability. *Sustainability* **2021**, *13*, 2038. [CrossRef]
10. Corazza, L.; Cottafava, D.; Torchia, D. Education for sustainable development: A critical reflexive discourse on a transformative learning activity for business students. *Environ. Dev. Sustain.* **2022**, 1–21. [CrossRef]
11. Menegat, R. Participatory democracy and sustainable development: Integrated urban environmental management in Porto Alegre, Brazil. *Environ. Urban.* **2002**, *14*, 181–206. [CrossRef]
12. Bazzoli, J.A. Agenda 2030: Extensão como trajeto para institucionalização. *Rev. Conex. UEPG* **2021**, *17*, 1–16. [CrossRef]
13. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised Cross-lingual Representation Learning at Scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 8440–8451. [CrossRef]
14. Li, X.; Tramèr, F.; Liang, P.; Hashimoto, T. Large Language Models Can Be Strong Differentially Private Learners. *arXiv* **2021**, arXiv:2110.05679. [CrossRef]
15. Makridakis, S.; Petropoulos, F.; Kang, Y. Large Language Models: Their Success and Impact. *Forecasting* **2023**, *5*, 536–549. [CrossRef]
16. Bodini, M. Opinion mining from machine translated Bangla reviews with stacked contractive auto-encoders. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *14*, 12119–12131. [CrossRef]
17. Bodini, M. Aspect Extraction from Bangla Reviews Through Stacked Auto-Encoders. *Data* **2019**, *4*, 121. [CrossRef]

18. Souza, F.; Nogueira, R.; Lotufo, R. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In Proceedings of the 9th Brazilian Conference on Intelligent Systems (BRACIS), Rio Grande, Brazil, 20–23 October 2020; pp. 403–417. [CrossRef]
19. Rodrigues, L.; Xavier, C.; Costa, N.; Batista, H.; Silva, L.F.B.; Chaleghi de Melo, W.; Gasevic, D.; Ferreira Mello, R. LLMs Performance in Answering Educational Questions in Brazilian Portuguese: A Preliminary Analysis on LLMs Potential to Support Diverse Educational Needs. In Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25), New York, NY, USA, 3–7 March 2025; pp. 865–871. [CrossRef]
20. Cortes, E.G.; Vianna, A.L.; Martins, M.; Rigo, S.; Kunst, R. LLMs and Translation: Different approaches to localization between Brazilian Portuguese and European Portuguese. In Proceedings of the 16th International Conference on Computational Processing of Portuguese, Santiago de Compostela, Spain, 14–15 March 2024; Volume 1, pp. 45–55.
21. Feretzakis, G.; Vagena, E.; Kalodanis, K.; Peristera, P.; Kalles, D.; Anastasiou, A. GDPR and Large Language Models: Technical and Legal Obstacles. *Future Internet* **2025**, *17*, 151. [CrossRef]
22. Almeida, R.; Amorim, E. A Legal Framework for Natural Language Model Training in Portugal. In Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024, Turin, Italy, 20 May 2024; pp. 6–12.
23. Nicolas, C.; Kim, J.; Chi, S. Natural language processing-based characterization of top-down communication in smart cities for enhancing citizen alignment. *Sustain. Cities Soc.* **2021**, *66*, 102674. [CrossRef]
24. Aleixo, A.M.; Azeiteiro, U.M.; Leal, S. Are the sustainable development goals being implemented in the Portuguese higher education formative offer? *Int. J. Sustain. High. Educ.* **2020**, *21*, 336–352. [CrossRef]
25. Serafini, P.G.; de Moura, J.M.; de Almeida, M.R.; de Rezende, J.F.D. Sustainable Development Goals in Higher Education Institutions: A systematic literature review. *J. Clean. Prod.* **2022**, *370*, 133473. [CrossRef]
26. Colglazier, E.W. Sustainable development agenda: 2030. *Science* **2015**, *349*, 1048–1050. [CrossRef] [PubMed]
27. Gratzer, G.; Muhar, A.; Winiwarter, V.; Lindenthal, T.; Radinger-Peer, V.; Melcher, A. The 2030 Agenda as a challenge to life sciences universities. *GAIA Ecol. Perspect. Sci. Soc.* **2019**, *28*, 100–105. [CrossRef]
28. Walsh, P.P.; Murphy, E.; Horan, D. The role of science, technology and innovation in the UN 2030 agenda. *Technol. Forecast. Soc. Chang.* **2020**, *154*, 119957. [CrossRef]
29. Colglazier, E.W. The sustainable development goals: Roadmaps to progress. *Sci. Dipl.* **2018**, *7*.
30. McCowan, T. Universities and the post-2015 development agenda: An analytical framework. *High. Educ.* **2016**, *72*, 505–523. [CrossRef]
31. Goralski, M.A.; Tan, T.K. Artificial intelligence and sustainable development. *Int. J. Manag. Educ.* **2020**, *18*, 100330. [CrossRef]
32. Bodini, M. Generative Artificial Intelligence and Regulations: Can We Plan a Resilient Journey Toward the Safe Application of Generative Artificial Intelligence? *Societies* **2024**, *14*, 268. [CrossRef]
33. Filho, W.L.; Sierra, J.; Price, E.; Eustachio, J.H.P.P.; Novikau, A.; Kirrane, M.; Dinis, M.A.P.; Salvia, A.L. The role of universities in accelerating the sustainable development goals in Europe. *Sci. Rep.* **2024**, *14*, 15464. [CrossRef]
34. Steele, W.; Rickards, L. *The Sustainable Development Goals in Higher Education: A Transformative Agenda?*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021. [CrossRef]
35. Mori Junior, R.; Fien, J.; Horne, R. Implementing the UN SDGs in Universities: Challenges, Opportunities, and Lessons Learned. *Sustain. J. Rec.* **2019**, *12*, 129–133. [CrossRef]
36. Chaleta, E.; Saraiva, M.; Leal, F.; Fialho, I.; Borralho, A. Higher education and Sustainable Development Goals (SDG)—Potential contribution of the undergraduate courses of the School of Social Sciences of the University of Évora. *Sustainability* **2021**, *13*, 1828. [CrossRef]
37. Blasco, N.; Brusca, I.; Labrador, M. Drivers for Universities' Contribution to the Sustainable Development Goals: An Analysis of Spanish Public Universities. *Sustainability* **2020**, *13*, 89. [CrossRef]
38. Ndubuka, N.N.; Rey-Marmonier, E. Capability approach for realising the Sustainable Development Goals through Responsible Management Education: The case of UK business school academics. *Int. J. Manag. Educ.* **2019**, *17*, 100319. [CrossRef]
39. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Fuso Nerini, F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* **2020**, *11*, 233. [CrossRef] [PubMed]
40. Matsui, T.; Suzuki, K.; Ando, K.; Kitai, Y.; Haga, C.; Masuhara, N.; Kawakubo, S. A natural language processing model for supporting sustainable development goals: Translating semantics, visualizing nexus, and connecting stakeholders. *Sustain. Sci.* **2022**, *17*, 969–985. [CrossRef]
41. Angin, M.; Taşdemir, B.; Yılmaz, C.A.; Demiralp, G.; Atay, M.; Angin, P.; Dikmener, G. A RoBERTa Approach for Automated Processing of Sustainability Reports. *Sustainability* **2022**, *14*, 16139. [CrossRef]
42. Firoozeh, N.; Nazarenko, A.; Alizon, F.; Daille, B. Keyword extraction: Issues and methods. *Nat. Lang. Eng.* **2020**, *26*, 259–291. [CrossRef]

43. Alqahtani, T.; Badreldin, H.A.; Alrashed, M.; Alshaya, A.I.; Alghamdi, S.S.; bin Saleh, K.; Alowais, S.A.; Alshaya, O.A.; Rahman, I.; Al Yami, M.S. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Res. Soc. Adm. Pharm.* **2023**, *19*, 1236–1242. [CrossRef]
44. Zhu, Q. Natural Language Processing in Teacher Training: A Systematic Review. *Lect. Notes Educ. Psychol. Public Media* **2023**, *18*, 83–90. [CrossRef]
45. Sousa, S.; Kern, R. How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing. *Artif. Intell. Rev.* **2022**, *56*, 1427–1492. [CrossRef]
46. Feretzakis, G.; Papaspyridis, K.; Gkoulalas-Divanis, A.; Verykios, V.S. Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review. *Information* **2024**, *15*, 697. [CrossRef]
47. Surabhi, M.C. Natural language processing future. In Proceedings of the 2013 International Conference on Optical Imaging Sensor and Security (ICOSS), Coimbatore, India, 2–3 July 2013; pp. 1–3. [CrossRef]
48. Hardeniya, N.; Perkins, J.; Chopra, D.; Joshi, N.; Mathur, I. *Natural Language Processing: Python and NLTK*; Packt Publishing Limited: Birmingham, UK, 2016; 702p.
49. Instituto Brasileiro de Geografia e Estatística (IBGE). Indicadores Brasileiros para os Objetivos de Desenvolvimento Sustentável. 2023. Available online: https://odsbrasil.gov.br/home/NewHome (accessed on 26 June 2023).
50. Lima, E.d.S.; de Oliveira, U.R.; Costa, M.d.C.; Fernandes, V.A.; Teodoro, P. Sustainability in Public Universities through lean evaluation and future improvement for administrative processes. *J. Clean. Prod.* **2023**, *382*, 135318. [CrossRef]
51. Klein, L.L.; De Guimarães, J.C.F.; Severo, E.A.; Dorion, E.C.H.; Schirmer Feltrin, T. Lean practices toward a balanced sustainability in higher education institutions: A Brazilian experience. *Int. J. Sustain. High. Educ.* **2021**, *24*, 259–278. [CrossRef]
52. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python*; O'Reilly: Sebastopol, CA, USA, 2009; 502p.
53. Bird, S. NLTK: The natural language toolkit. In Proceedings of the COLING/ACL on Interactive Presentation Sessions (COLING-ACL '06), Sydney, Australia, 17–18 July 2006; pp. 69–72. [CrossRef]
54. Flores, F.N.; Moreira, V.P.; Heuser, C.A. Assessing the impact of stemming accuracy on information retrieval. In *Computational Processing of the Portuguese Language, Proceedings of the 9th International Conference, Porto Alegre, Brazil, 27–30 April 2010*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 11–20.
55. Dusetzina, S.B.; Tyree, S.; Meyer, A.M.; Meyer, A.; Green, L.; Carpenter, W.R. An overview of record linkage methods. In *Linking Data for Health Services Research: A Framework and Instructional Guide [Internet]*; Agency for Healthcare Research and Quality: Rockville, MD, USA, 2014.
56. Yang, A.; Zhang, J.; Pan, L.; Xiang, Y. Enhanced twitter sentiment analysis by using feature selection and combination. In Proceedings of the 2015 International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec), Hangzhou, China, 16–18 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 52–57.
57. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (SciPy), Austin, TX, USA, 18 June–3 July 2010; pp. 56–61. [CrossRef]
58. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef] [PubMed]
59. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
60. Waskom, M.L. seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
61. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]
62. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
63. Farrokhnia, M.R.; Banihashem, S.K.; Noroozi, O.; Wals, A.E.J. A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research. *Innov. Educ. Teach. Int.* **2023**, *61*, 460–474. [CrossRef]
64. Gursoy Haksevenler, B.H.; Kavak, F.F.; Akpinar, A. Separate waste collection in higher education institutions with its technical and social aspects: A case study for a university campus. *J. Clean. Prod.* **2022**, *367*, 133022. [CrossRef]
65. Jakimiuk, A.; Matsui, Y.; Podlasek, A.; Koda, E.; Goli, V.S.N.S.; Voběrková, S.; Singh, D.N.; Vaverková, M.D. Closing the loop: A case study on pathways for promoting sustainable waste management on university campuses. *Sci. Total. Environ.* **2023**, *892*, 164349. [CrossRef]
66. Cottafava, D.; Ascione, G.S.; Corazza, L.; Dhir, A. Sustainable development goals research in higher education institutions: An interdisciplinarity assessment through an entropy-based indicator. *J. Bus. Res.* **2022**, *151*, 138–155. [CrossRef]
67. Wals, A.E. Sustainability in higher education in the context of the UN DESD: A review of learning and institutionalization processes. *J. Clean. Prod.* **2014**, *62*, 8–15. [CrossRef]
68. Braßler, M. Interdisciplinary Problem-Based Learning—A Student-Centered Pedagogy to Teach Social Sustainable Development in Higher Education. In *Teaching Education for Sustainable Development at University Level*; Leal Filho, W., Pace, P., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 245–257. [CrossRef]

69. Sterling, S. Education in change. In *Education for Sustainability*; Routledge: London, UK, 2014; pp. 18–39.
70. Lidgren, A.; Rodhe, H.; Huisingh, D. A systemic approach to incorporate sustainability into university courses and curricula. *J. Clean. Prod.* **2006**, *14*, 797–809. [CrossRef]
71. Viera Trevisan, L.; Leal Filho, W.; Ávila Pedrozo, E. Transformative organisational learning for sustainability in higher education: A literature review and an international multi-case study. *J. Clean. Prod.* **2024**, *447*, 141634. [CrossRef]
72. Agbedahin, A.V. Sustainable development, Education for Sustainable Development, and the 2030 Agenda for Sustainable Development: Emergence, efficacy, eminence, and future. *Sustain. Dev.* **2019**, *27*, 669–680. [CrossRef]
73. Körfgen, A.; Förster, K.; Glatz, I.; Maier, S.; Becsi, B.; Meyer, A.; Kromp-Kolb, H.; Stötter, J. It's a Hit! Mapping Austrian Research Contributions to the Sustainable Development Goals. *Sustainability* **2018**, *10*, 3295. [CrossRef]
74. Annan-Diab, F.; Molinari, C. Interdisciplinarity: Practical approach to advancing education for sustainability and for the Sustainable Development Goals. *Int. J. Manag. Educ.* **2017**, *15*, 73–83. [CrossRef]
75. Eagan, P.; Cook, T.; Joeres, E. Teaching the importance of culture and interdisciplinary education for sustainable development. *Int. J. Sustain. High. Educ.* **2002**, *3*, 48–66. [CrossRef]
76. Zhou, L.; Rudhumbu, N.; Shumba, J.; Olumide, A. Role of Higher Education Institutions in the Implementation of Sustainable Development Goals. In *Sustainable Development Goals and Institutions of Higher Education*; Nhamo, G., Mjimba, V., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 87–96. [CrossRef]
77. Waas, T.; Verbruggen, A.; Wright, T. University research for sustainable development: Definition and characteristics explored. *J. Clean. Prod.* **2010**, *18*, 629–636. [CrossRef]
78. DeFries, R.S.; Ellis, E.C.; Chapin III, F.S.; Matson, P.A.; Turner, B.; Agrawal, A.; Crutzen, P.J.; Field, C.; Gleick, P.; Kareiva, P.M.; et al. Planetary opportunities: A social contract for global change science to contribute to a sustainable future. *BioScience* **2012**, *62*, 603–606. [CrossRef]
79. Clark, S.G.; Wallace, R.L. Integration and interdisciplinarity: Concepts, frameworks, and education. *Policy Sci.* **2015**, *48*, 233–255. [CrossRef]
80. Leal Filho, W. Teaching sustainable development at university level: Current trends and future needs. *J. Balt. Sci. Educ.* **2010**, *9*, 273–284.
81. Cobianchi, L.; Dal Mas, F.; Barcellini, A.; Vitolo, V.; Facoetti, A.; Peloso, A.; Massaro, M.; Vanoli, A.; Brugnatelli, S.; Ciocca, M.; et al. Knowledge translation in challenging healthcare environments: The PIOPPO experience at the National Centre of Oncological Hadrontherapy (CNAO Foundation). In Proceedings of the 21st European Conference on Knowledge Management, ECKM, Online, 2–4 December 2020; p. 124.
82. Cobianchi, L.; Dal Mas, F.; Angelos, P. One Size Does Not Fit All—Translating Knowledge to Bridge the Gaps to Diversity and Inclusion of Surgical Teams. *Ann. Surg.* **2021**, *273*, e34–e36. [CrossRef] [PubMed]
83. Summers, M.; Childs, A.; Corney, G. Education for sustainable development in initial teacher training: Issues for interdisciplinary collaboration. *Environ. Educ. Res.* **2005**, *11*, 623–647. [CrossRef]
84. Woiwode, H.; Froese, A. Two hearts beating in a research centers' chest: How scholars in interdisciplinary research settings cope with monodisciplinary deep structures. *Stud. High. Educ.* **2021**, *46*, 2230–2244. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.