Applying AI for social good: Aligning academic journal ratings with the United Nations Sustainable Development Goals (SDGs)

David Steingard, Director, SDG Dashboard, Associate Professor of Management

Marcello Balduccini, SDG Dashboard Fellow, Assistant Professor of Decision and System Sciences

Akanksha Sinha, SDG Dashboard Data Scientist

Saint Joseph's University
5600 City Avenue, Philadelphia, PA USA
{steingar, mbalducc, akanksha.sinha}@sju.edu

Abstract

This paper offers three contributions to the burgeoning movements of AI for Social Good (AI4SG) and AI and the United Nations Sustainable Development Goals (SDGs). First, we introduce the SDG-Intense Evaluation framework (SDGIE) that aims to situate variegated automated/AI models in a larger ecosystem of computational approaches to advance the SDGs. In order to foster knowledge collaboration for solving complex social and environmental problems encompassed by the SDGs, the SDGIE framework details a benchmark structure of data-algorithm-output to effectively standardize AI approaches to the SDGs. Second, as a specific instantiation of the SDGIE framework, the SDG Impact Intensity Model (SDGIIM) is theoretically and operationally established. SDGIIM embeds expert decision-making and SDG keyword banks in textual data processing to determine overall SDG "impact intensity." Ideally, SDGIIM can be applied to textual data sets from any sector or discipline: academia, business, government, non-profit, civil society, etc. Third, the SDGIIM instantiation is applied to the specific domain of academic journal rating systems as a case study. Traditionally, academic journals have been evaluated on loosely conceived and empirically shaky notions of 'quality.' Aligned with the trend of AI4SG and broader calls to action, 'impact' is rapidly becoming the primary normative consideration for assessing academic journals. We hypothesize and demonstrate that SDGIIM is capable of producing evaluations aligned with experts' expectations of SDG impact intensity; the consistent analysis and rating of textual data sets that embody the SDGs with varying degrees of meaning and, ultimately, promote positive impact on the actual material conditions of the world.

Keywords
Sustainable Development Goals; SDG Impact Intensity; AI for social good (AI4SG); Sustainable AI; Computational Sustainability; journal ratings

Acknowledgments
Portions of this publication and research effort are made possible through the financial support of the Johnson & Johnson Foundation. We thank Simon Linacre and Cabells for our fruitful collaboration on the development and deployment of this research.

Declarations
Funding: see acknowledgments section.
Conflicts of interest/Competing interests: none.
Availability of data and material: N/A.
Code availability: N/A.
Ethics approval: N/A.
Consent to participate: N/A.
Consent for publication: N/A.

Table of Contents

1. Introduction: SDGs and AI for social good

The 17 United Nations Sustainable Development Goals or SDGs (United Nations, webpage-b; see Figure 1) and accompanying Agenda 2030 (United Nations, 2015) represents a bold and detailed plan to deliver human flourishing and environmental sustainability by 2030. This plan requires concerted global multi stakeholder collaboration across all sectors of society—government, business, civil society, non-governmental organizations, and, as the focus of this paper, academia.

Fig. 1 The 17 United Nations Sustainable Development Goals (SDGs)

Through the use of AI evaluation methods, this paper examines how the primary engine of academia, published journal research, contributes to the alignment and advancement of the SDGs by using automated — and in particular AI-based — techniques. Our aim in this paper is framed within the clear edict of "Sustainable AI'' offered by (Vinuesa et al., 2020):

> Specifically, the aim is to understand whether this branch of computer science can influence production and consumption patterns to achieve sustainable resource management according to Sustainable Development Goals (SDGs) outlined in the UN 2030 Agenda. (7)

We offer a contribution to the burgeoning field of AI and ethics focusing on AI for 'social good' (Google AI, webpage) or AI4SG (Floridi et al., 2020; Purdy, 2020; Tomašev et al., 2020; Chui et al., 2019; Gomes, 2019; AISOC, 2017; AAAI, 2017) and specifically, the subdomain of AI and the SDGs (Cowls et al., 2021; Di Vaio et al., 2020; Goralski and Tan, 2020; Hager et al., 2019; Kriebitz and Lütge, 2020: 85; Tsamados et al., 2021: 12; and Vinuesa et al., 2020). Our aim is to contribute insight to this exciting movement, deploying AI insights to advance the SDGs in order to tackle some of the world's most exigent human and environmental problems.

What we call SDG-Intense Evaluation, or SDGIE (see Figure 2), contextualizes how various approaches to SDG-related evaluations constitute a framework to undergird discussion throughout the paper. We offer the SDGIE framework as an organizing schema encompassing various approaches to evaluating how artificial intelligence assesses SDG-related contributions of AI models and applications. SDGIE supports (Cowls et al., 2021) insightful observation that AI4SG as a field embodies a "lack of normative analyses and a shortage of empirical evidence" and "advocates the use of the United Nations' Sustainable Development Goals (SDGs) as a benchmark for tracing the scope and spread of AI4SG" (p.1).

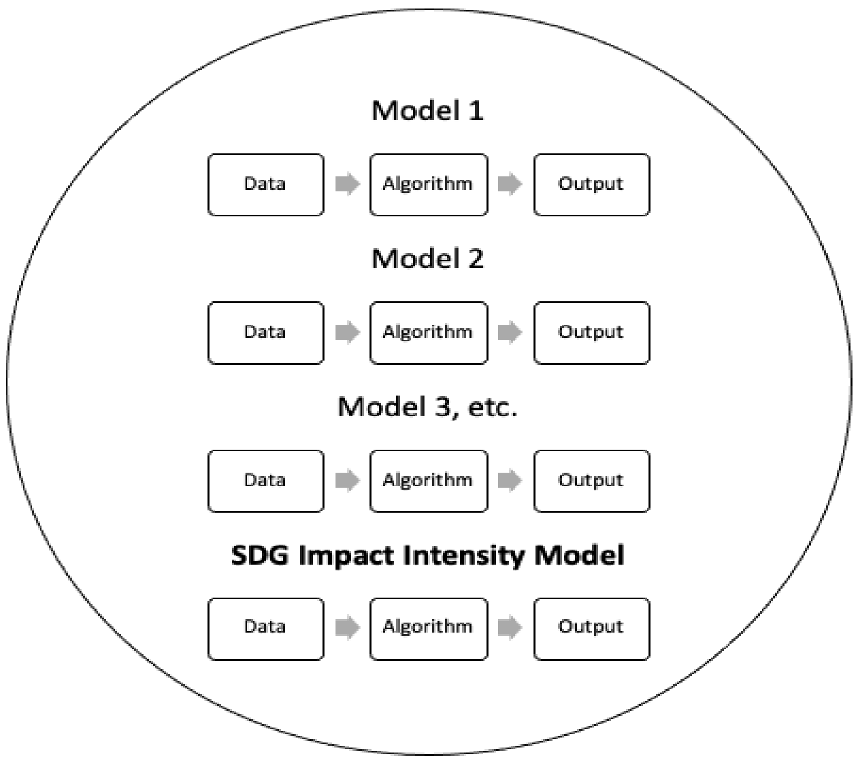

Fig. 2 SDG Impact Evaluation (SDGIE) Framework

The SDG Impact Intensity Model, or SDGIIM, as described hereafter, is one instantiation of possible approaches in the SDGIE. By framing a larger ecosystem for these models, we establish a domain in which approaches can be compared, discriminated, and ultimately improved upon. Iterating upon the efficacy of each approach enables expansive facilitation for identifying content with SDG-intensity and contributes directly to the acceleration of SDG realization, regardless of the application.

It should be noted that what ‘counts’ as content with SDG-intensity is fundamentally an issue of normative ethical standards, subjectively determined by a variety of human inputs and culture. Given the use of an algorithm to adjudicate standards in SDGIE, this paper will carefully examine the expert human intelligence and decision-making that inform the algorithm. As Véliz (2021: 1) instructs, “algorithms are not moral agents,” so moral agency derives from how an algorithm is programmed by humans. It is only by fully understanding the moral logic, normative assumptions, and critical discernment powering an algorithm that ‘social good’ can be adjudicated. Our aim is to reveal the key challenges of SDGIE in order to continually improve its constituent instantiation models and deliver maximum impact on the SDGs.

It is necessary to provide additional background on how our approach fulfils the social good as the AI involved in SDGIIM is a second-order, yet very potent determinant of advancing or inhibiting ‘social good’ in academic publishing, academia, and ultimately, the material circumstances of the world. Conventionally, positive or negative impacts of AI are considered first-order in nature. AI is utilized to generate actionable data and insights that directly affect human and environmental outcomes. Algorithm outputs can have beneficent effects as seen in myriad contexts. Predictive analytics generate accurate weather predictions. Supply chain AI helps maintain the vital flow of goods and services. AI in genomic analytics aids advancements in the development of medicines and treatments. Cybersecurity AI provides for personal, military, and industrial security and continuity. Of course, AI’s impacts can foster deleterious social and environmental consequences as well. AI has been deployed for nefarious ends of authoritarian cyber police states, “surveillance capitalism” (Wylie et al., 2022), racial and gender profiling, and fomenting an addictive and injurious “attention economy” through social media (Bhargava & Velasquez, 2021).

For the purposes of our paper, we locate the moral impact of academic journals in relationship to AI as a twofold second-order effect. First, academic journal articles are consumed for educational, policy, scientific, managerial, etc. purposes. Knowledge from academic journals is employed in all aspects of human existence and significantly shapes behavior and outcomes, both positively and negatively. Admittedly, this causal chain of the written words into actions (Steingard & Linacre, in press) is not directly a result of algorithm outputs or AI technique; it is not first-order causality. However, it is precisely sophisticated AI approaches that power how academic journals are valued in terms of their quality and, presumably, demonstrable impact in the world.

In short, AI determines what academic journals are ‘better’ than others by criteria fed to algorithms primarily focused on counting journal article citations–”citation-intensity” (Steingard & Linacre, 2021)--rather than the SDG Impact Intensity we are offering in this paper. For example, Google Scholar’s (Google, webpage) algorithm’s primary determines its “Top publications” by using the “h5-index” which is a purely quantitative counter of citations with no qualitative evaluation of impact included. Contrastingly, myriad search strings and portals for finding SDGs in academic journals exist (Purnell, 2022). Bibliometric outputs are solely determined by the underlying, second-order, criteria built-in to their algorithms. There is no amoral position for these algorithms–they either count quantitatively with no indication of impact or they specifically analyze content to ascertain alignment with normative standards of positive impact as embodied in, for the purposes of this paper, the SDGs. Shifting the paradigm of academic journal evaluation from counting to impact requires a fundamental paradigm shift in the algorithms that power ratings, search tools, bibliometrics, and other analytics that determine their value. Thus, SDGII, being powered by AI in this second-order modality, adds squarely to the AI for ‘social good’ called for in this special issue.

The remainder of this paper examines the application of one particular model of SDGIE—the SDG Impact Intensity Model or SDGIIM, that analyzes academic research impact in alignment with the SDGs. Our approach formalizes the notion of impact, and allows AI to operationalize that impact—effectively marrying the positive impacts outlined by the SDGs with a more principled approach to apply AI for “‘social good,’ enabling the deployment of revolutionary services'' that “will meaningfully impact societal development and sustainability” (AI & SOCIETY, 2021: 1). We hope our SDG-centered contributions in this paper advance this inspiring call to action.

2. Methodology

2.1 SDGs in academic publishing

Examining our particular model of the SDGIE framework—the SDG Impact Intensity Model (SDGIIM), whose development was inspired by research we previously conducted on reporting and best-practice sharing of activities related to the SDGs in our SDG Dashboard (Garwood et al., 2020). In this section we apply our SDG Impact Intensity model to an application involving the rating of academic journals in a collaborative partnership between Saint Joseph’s University and Cabells Scholarly Analytics (Linacre, 2021a,) whose mission is: “To provide academics with accurate information and reputable outlets for publication” (Cabells, webpage-a: 1).

We chose our analysis of SDG-intense evaluation in academic publishing because of an upswell of activity in this sector. Since ratification by 193 countries in 2015, academic interest in publishing about the SDGs continues to increase (Nakamura et al., 2019). Evidence for this claim is provided by examining the sheer number of search engines and publisher portals that allow academic researchers to query academic research content through the lens of the SDGs. A review of these SDG gateways produces a number of open-source and commercial resources to query academic content and research vis-à-vis the SDGs: Digital Science & Research Solutions, webpage; Elsevier, webpage; Taylor & Francis, webpage; Springer Nature, webpage; RELX plc, webpage; Rotterdam School of Management, webpage; ScienceOpen, webpage; Emerald Publishing, webpage; Linacre, 2021b; Responsible Research in Business Management[1]). Additionally, as an important recent development through the United Nations, the SDG Publishers Compact now serves as an official signatory organization "designed to inspire action among publishers...to accelerate progress towards the Sustainable Development Goals (SDGs) by 2030" (United Nations, 2015: 1).

Common to all of these gateways are algorithms that adjudicate what 'counts' as academic research supporting the SDGs—a quest to find construct validity around the idea of actual social impact (Bornmann et al., 2019; Ravenscroft et. al., 2017). However, the underlying standards, methodology, techniques, etc. behind these approaches are largely opaque, and for good reasons. While collaboration on advancing the accuracy of the SDGIE framework is possible, there are justifiably competitive and market-based reasons why different providers would not open-source their proprietary code or normative decision-rules underlying the algorithms. However, by establishing the SDGIE framework for understanding different approaches to computing SDG impact, there is now common ground for mutual investigation and development of different SDGIE models generally.[2]

We begin with the question: *How do we ascertain if a particular academic journal accurately reflects an embodiment of the SDGs and engenders impact?* (Jack, 2020). What would make a journal SDG-affirming, neutral, or possibly even contravene impact on the SDGs? Two key demands arise from these questions. First, we must understand the substantive essence of the SDGs. What do the SDGs mean and how do we interpret their meaning? Second, with an understanding of the SDGs, how do we apply this understanding to our academic journal unit of analysis? How can we measure to what extent a journal embodies the SDGs—how SDG-intense is that journal? Moreover, SDG Impact Intensity adds another demand in terms of the type of meaning required—impact, the degree to which any particular journal not

[1] "UN SDG related issues for management" is a sub-gateway of the more encompassing Responsible Management Gateway.

[2] To this end, and in the spirit of SDG #17: Partnerships for the Goals, the authors are collaborating with researchers from Guelph University's Lang School of Business (Rodenburg et al., 2021) and the Rotterdam School of Management (webpage), exchanging ideas and testing outcomes for convergent/discriminant validity of our respective techniques.

only reflects the SDGs to some degree, but also offers scholarly theories, constructs, applications, cases, tools, etc. that can arguably be deemed to advance the SDGs' core purpose and material manifestation. Does a journal accelerate (or possibly deter) the fulfillment of particular SDGs that improve the human condition and promote environmental sustainability in the world?

The ultimate objective of the SDGIIM academic journal rating system is to provide an accurate evaluation of how much SDG-intense content is present in an academic journal, our chosen domain for application. While the scope of application of the rating system is limited in this paper to academic journals as the unit of analysis, it is plausible that the SDGIIM could be applied to other domains that involve representations of SDGs in text, e.g., corporate reports (Kulevicz et al., 2020), government reports, policies, websites, books, project descriptions, individual journal articles, social media, databases, etc. Common to all of these domains is the analysis of textual expressions; language as it occurs naturally in a domain that is interrogated and processed via the SDGIIM model's algorithm.

2.2 SDGIIM instantiation

In this section we discuss in more detailed terms the motivations for the choices made in the instantiation of the SDGIIM; an outcome metric that assesses the degree to which academic journals reflect SDGs in their published articles. This metric is derived from consideration of a number of possible assessment options. What does it mean for a journal and its articles to be "intense" in terms of SDGs? Intensity as we define it is a combination of three primary factors. First, SDGIIM appraises the *meaningfulness* of the relationship between the content of the article and the SDGs. How central are the SDGs to the thesis or purpose of the article? For example, an article that analyzes and makes recommendations about promoting gender equality in the workplace would be directly supportive of SDG#5: Gender Equality. This article, *prima facie,* would be cast as very intense vis-à-vis this particular SDG. Additionally, the article would definitely have some high intensity around SDG#10: Reduced Inequalities. Conversely, an article that promoted business strategies to capitalize on industrial production in the developing world where worker and environmental standards are lax, and even exploitive, would receive a very low intensity rating. Meaning here is adjudicated by the actual language used in the aggregate SDG keywords applied to the SDGs and its Targets. For this factor of meaning, journal article intensity is determined by the word matching generated from the algorithm—to what degree do we find SDG keywords in the article?

We compiled a list of 1,095 keywords that we collected to represent the SDGs. We obtained meaningful keywords from RELX (webpage) and SDSN Australia, New Zealand and

Pacific (2021). We also utilized The Big Benchmarking Tool (University of Worcester & Kingston University London, webpage) to organize these keywords into sub-lists for the specific SDGs they represent.

The SDGIIM relies on *frequency* as one of its factors. Frequency provides a measure of how many times SDG keywords occur in a journal. Continuing with the example from SDG#5: Gender Equality, core keywords like "women(s')"; "empowerment"; and "equal opportunities" would occur multiple times in related articles. Of course, syntax and word strings matter—e.g., just discussing "empowerment" without a direct connection to "women" would not align. Later in the paper we discuss next steps to improve the algorithm to identify particular phrases and word combinations in context as meaningful. The factor of frequency is also further codified by including an analysis of *clustering*—different groups of keywords that demonstrate high frequency. It would be reasonable in a SDG#5: Gender Equality high intensity article to discover clusters of word groups. Clustering also adds a form of triangulation to the algorithm; the more words from the SDG keyword bank that align, the higher the probability that the article is actually reflective of the SDGs. If only the word "women" appeared in an article, that article could be about a variety of topics related or unrelated to SDG#5. This more sophisticated type of *frequency* and *clustering* analyses are critical to the success of the algorithm to determine the accuracy of the SDGIIM.

In addition to *meaning, frequency,* and *clustering,* another factor that undergirds SDG SDGIIM is *weighting* in the algorithm. Weighting entails the assignment of multiplicative values for each of the words in the SDG keyword bank. The keyword bank is subdivided into three general categories. Category 0 contains all of the SDG keywords sourced from the most intense sources of keywords related to the SDGs—the actual wording of the SDGs using the language of the SDGs themselves (e.g., "SDG(s)," "sustainable development goals," "global goals") as well as certain keywords extracted from Agenda 2030 (e.g., "Agenda 2030" and "United Nations"). Categories 1 - 17 contain specific language from each of the Goals and accompanying Targets. Categories 18+ are a collection of expert recommended keywords that allow for recalibrations—human learning—of the algorithm to include more keywords that include SDG-intense journals that did not score well due to their technical language differentiation. These journals are certainly worthy of a solid SDG Impact Intensity rating, but effectively missed by the algorithm.

In our approach, keywords are the pillars of rating and categorization of journals. Word cloud analysis is another technique (Heimerl et al., 2014) frequently used in the context of keyword extraction and in the area of social and environmental impact akin to SDG-intense evaluation (Kulevicz et al., 2020). We have experimented with word clouds as a supplemental lens to evaluate and validate the results produced by the SDGIIM. A word cloud visually represents the frequency and importance of a set of words by varying the font size of the words. Creating and analyzing the world clouds produced for the journals of interest is an example of data exploration in which we form hypotheses about our defined problem or project goal (here assessing SDG Impact Intensity) by visually analyzing the data. In order to make the word clouds more meaningful, we refined the set of words by means of a preprocessing phase in which we removed stop words, numbers, non-English words, special characters, and applied lemmatization (Manning et al., 2008).

The final aspect of the model that must be instantiated is the measurement of the impact, i.e. of the degree to which any particular journal can be deemed to advance the SDGs core purpose and material manifestation of 'social good' in the world. For this purpose, the keywords are assigned weights, which are aimed at capturing how much a certain keyword may be evidence of impact. Because our aim is to assign weights that enable a transparent evaluation of impact, the current weighting scheme is intentionally minimalistic. The keywords of SDG_i and SDG_18 have a weight of 1. The keywords of SDG_0 have a weight of 17. This weighting scheme is designed to make the total weight from all of the SDG_i and SDG_18 equal to that of the SDG_0 keywords, in line with the view discussed earlier of the SDG_0 keywords as critical indicators representing the core SDG values and terms, unlike some of the SDG_i keywords, which may be ambiguous.

This tripartite keyword bank subdivision affords us the opportunity to fine tune how SDG Impact Intensity is derived in the SDGIIM. How heavily in the ultimate rating should each of these weigh in the final calculation? Continuing with the SDG#5 Gender Equality journal example to explicate weighting, this journal would obviously score highly on SDG #5. Yet, if it does go deeply on this one SDG and does not register on any other SDGs or is not explicitly contributing to Category 0, how weighty should its assignment be? Does a journal's depth on one SDG mean a higher SDG Impact Intensity score than another journal who less intensely addresses three SDGs? Does a journal that scores highly on Category 0—a direct hit for SDGs—but does not score well on any particular SDGs—count less than a journal that goes deep on one SDG? Calibrating the SDG score *weights* in concert with *meaning, frequency, and clustering,* evidences the biggest challenge in determining SDG Impact Intensity. Which words, combinations of words, and frequency of words define a valid and methodologically defensible attribution of SDG Impact Intensity to a journal?

Thus, we can begin to envision how an AI technique can assist in ascertaining a loosely or even unsubstantiated approach to "impact" with the model of the SDGIE framework, moving toward "intensity." Textual analysis via keyword banks will reveal a *rich milieu of meaning*—

keywords reflecting the SDGs present in the text ground the analysis. The *quantity of occurrences* provides another layer of discrimination. While sheer numbers do not necessarily indicate impact intensity, quantity is a vital dimension if tempered with another element in calculating intensity. *Magnitude* of keyword occurrences crosscuts a simple count of quantity by weighting each keyword in the overall calculation. Even with words common to the SDGs, there are particular words that reflect the deeper essence or particular facets of the SDGs as a composite system or individually as 17 dimensions. In the next section we detail how our technique incorporates these evaluative criteria to produce a valid construct of "SDG Impact Intensity" through the SDGIIM.

3. Case Study: SDGs in Academic Publishing

The emerging importance of SDGs in academic publishing makes for a practicable and fruitful domain in which to articulate, apply, and test the SDGIIM. Figure 3 combines the domain specific case application within the overall model, examining academic journals, as the unit of analysis, for degrees of SDG impact intensity. As mentioned earlier, these data sources are interchangeable with textual data from a variety of sectors (e.g., business, government, etc.). Our particular instantiation of this model can serve other models in the SDGIE framework as a benchmark for comparing, contrasting, and integrating best practices. The model is fundamentally an iterative, organic process that integrates both expert input and automated computation (e.g., automated reasoning and machine learning techniques) to support its evolution.

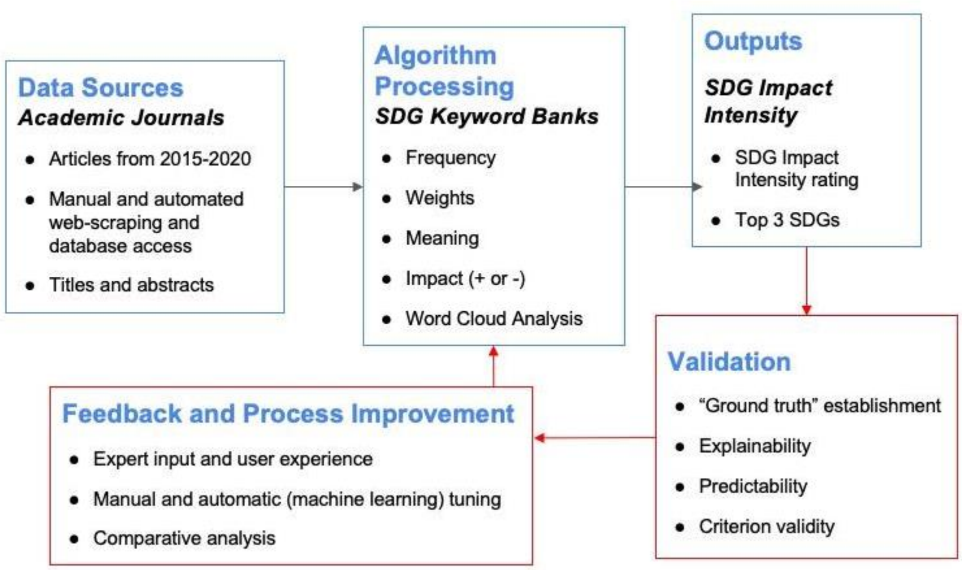

Fig. 3 SDG Impact Intensity Model (SDGIIM)

3.1 How the SDGIIM assesses SDG impact

Endemic to the SDG Impact Intensity model is the construct of SDG impact—capturing the substantive essence of the SDGs, i.e. what the SDGs mean and how one should interpret their meaning. As previously discussed, measuring the extent to which a journal embodies the SDGs is reduced to calculating the frequency and weights attributed to the keywords found throughout the titles and abstracts of the journal. Note that in this context, a keyword may be a multi-word phrase. We leverage various techniques to detect the occurrence of keywords in spite of morphological variations that do not affect the meaning and of the presence of additional words. A detailed discussion is omitted as it is outside the scope of the present paper.

It is important to discuss, more generally, what 'impact' means as an evaluative lens on academic research. With the recent development of advanced web scraping and big data analytics, academic publishers and institutions alike now employ myriad citation indices, social media indices, rankings, ratings, etc. to effectively adjudicate estimations of impact for academic research outputs. The UK has even made such evaluations integral to its national focus on "assessing the quality of research in UK higher education institutions" (REF, n.-d.: 1). Perhaps before examining the impact of academic outputs, it is prudent to consider 'alignment,' and in this particular case of SDGIE, alignment within the SDG gateways discussed earlier.

How do these portals discover matches between content and SDGs? Although all of these portals provide some overview of their methodologies, it is difficult to pinpoint the underlying models and applications that power them. Given the pure scale of these undertakings, all of these approaches employ some type of principled algorithmic technique that interrogates data, drawing conclusions about how well the content matches SDGs. Once alignment is achieved, challenging questions arise as to how this alignment fosters, and to what degree it impacts the world. Incorporation of SDG language in topics and applications found in academic research does not necessarily guarantee impact; yet discovering patterns of language and expressions is a defensible proxy for considering SDG iImpact Intensity.

Core to these approaches are a few underlying and oftentimes tacit assumptions about what impact means. Impact is usually associated with quantitative metrics that equate 'more' with ‘better’: the number of times a particular piece of published academic work is cited in other academic publications (e.g., SSCI, Google Scholar) and in social media (e.g., Altmetrics). Of course, a paper cited thousands of times versus one that receives a handful of citations is undoubtedly more well known and perhaps even influential, but does it generate the type of impact necessary to advance the SDGs with tangible results? There are currently scant mechanisms for interrogating how quantitative occurrences of academic citations correlate (or do not) with a clearly articulated, transparent, and widely accepted definition of impact. As we will see later, our particular application of the SDGIIM enriches this quest for maximum occurrences of data with specific, quantitative decision-rules and standards that provide a more robust and holistic approach to the assessment of impact .

To address this question of impact, we must first consider the underlying and hidden assumptions about what makes for “quality” (Aksnes et al., 2019) academic research. Traditionally, quality academic research is characterized by a number of factors: its theoretical and methodological rigors; peer-review assessment; citation counting; editorial board affiliations and reputations; social media mentions; journal rankings, etc. These aspects of quality do provide a solid foundation for academic research in terms of a standardized scientific approach that is vetted and legitimized as meaningful knowledge. Yet, there are no provisions for quality research to necessarily provide insights about positive or negative impacts on people and the planet. In other words, quality research is not readily equivalent with research that makes an impact on ethical and sustainable outcomes. It is also the case that by having self-reinforcing, closed systems of evaluating quality, there forms an elitist cabal of ‘top’ journals that are unchanging over time (Harley & Fleming, 2021). Once a journal is considered high quality based on subjective cultural approval—not adjudicated by objective scientific or ethical standards—there is a self-reinforcing, entrenched dynamic that maintains a subset of academic journals consistently at the apex of rating and rankings. It may be argued that this is the inverse of a virtuous circle.

No matter the actual impact of a journal, the fact that it is cited frequently and popular (not all that is popular is beneficial) in social media enables ranking entrenchment; becoming even that much more difficult for the ‘best’ journals to be unseated. And, this dynamic makes it extremely difficult for a new journal to break into the top echelons of journals that are considered high quality or, specifically in this paper, to make an impact. One could argue that low citation scores and social media hits indicate that a journal is demonstrably not meritorious of high esteem in academic circles. Yet, if these lower-rated, alternative journals do indeed deliver more demonstrable impact than highly rated ones, they should be considered with a different set of metrics and standards; this is precisely why we are introducing our alternative SDGIIM.

Moreover, it is arguably the case that many of the vexing social and environmental challenges of today have manifested because "quality" research (Pontille, D., & Torny, 2010) is not always conducive to the publication of valuable research that impacts positive results. Moreover, the attendant analytical techniques supporting notions of quality are calibrated and biased to produce metrics of quality that may be detached from impact. Traditional notions of quality journals include objective data of acceptance rates, composition of editorial boards, and quantification of citation hits. Additionally, there are other subjective data like deans' surveys, social media analytics (Bornmann et al, 2019), and published journal rankings (Purnell, 2022) that factor into evaluative schema of *quality.* However, these inputs do not necessarily provide insights as to the *impact* journals may have on ethical and sustainable outcomes for humanity and the Earth. Arguably, given the range of troubling global issues like climate change, poverty, economic injustice, ecosystem destruction, human rights violations, etc., it may be the case that quality journals—by not directly addressing remedies for these issues directly—in fact contribute indirectly to these deleterious issues, creating negative impacts.

By way of illustration, a particularly well-received paper does not necessarily guarantee positive (or avoid negative) impact on person and planet. For instance, leading scholarship on mega-scaling fossil fuel extraction economics and practices—while perhaps cutting-edge in a conventional energy paradigm—is arguably generating a negative impact on the Earth due to its associated greenhouse gas emissions that foster global warming and climate change. Thus, given how traditional academic cultures and standards produce top-rated academic scholarship of impact based primarily on quantitative metrics and self-referencing, it is plausible to postulate that not all impact ascribed to academic scholarship is for the 'good' no matter how well intended.

There exist particular academic cultural biases in determining impact. Harley and Fleming (2021: 1) found that only 2.8% of elite management journal articles between 2008 and 2010 "addressed global 'grand challenges'--such as inequality, climate change, racism, and gender discrimination." Journal rankings in academic disciplines are usually determined by surveys of academics and academic administrators without consideration of any consistent normative ethical standards related to impact. It would be unfounded to claim these cultural productions of academic publishing impact to be mere popularity contests. However, the guiding principle for determining impact by academics oftentimes has more to do with intra-academic standards of impact than with the extra-academic contributions of academic research to the greater good of society or the sustainability of the Earth (Responsible Research in Business Management, webpage) .

Compounding this self-referential dynamic is the dominance of particular publishers as bellwethers of impact for a particular field. To advance the conversation about how impact is determined in academic publishing, we offer a conceptualization and test of a model for an

SDGIE evaluation schema that incorporates directly the normative ethical imperatives of human rights, social, gender, and racial justice, environmental sustainability, economic prosperity, and peace grounded in the SDGs. Utilizing SDGs as an ethical undergirding for assessing the positive human and environmental impacts of academic research is the first step to reframing impact. Fortunately, there is a great deal of positive momentum as reflected by a demonstrable commitment by the major leaders of academic publishing as signatories of the SDG Publishers Compact (United Nations, 2015).

3.2 Application of SDGIIM to academic journal publishing data

As for data, the SDGIIM was applied to a subset of the approximately 3,000 business and business related journals listed on Cabells: 50 journals from the Financial Times FT50 (Ormans, 2016) and 50 handpicked (by the staff at Cabells and Saint Joseph's University), expert-evaluated journals presumed to be particularly rich with SDG impact. These two types of journals were chosen purposefully as a method to test whether or not the SDGIIM would be capable of accurately differentiating extreme polarities of SDG Impact Intensity and even more refined gradations—*effectively, this is the working hypothesis depicted in Figure 4.* More detail on the datasets follows below.

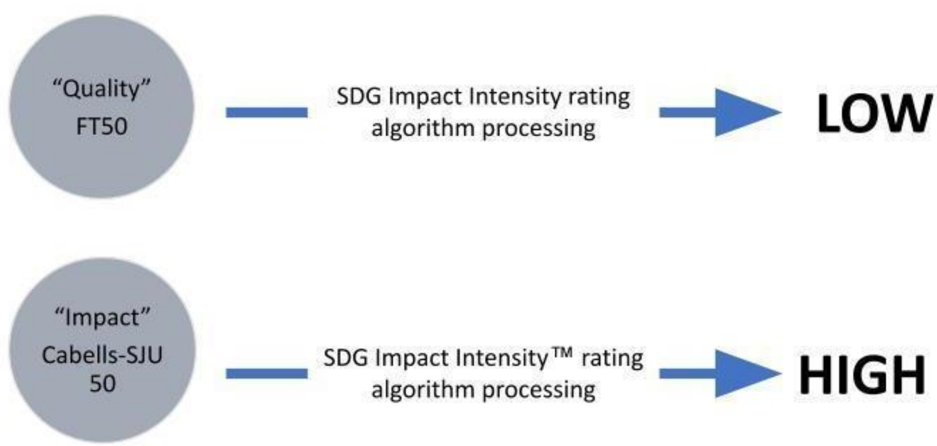

Fig. 4 Predictive Hypothesis for SDG Impact Intensity

It is worth remarking that FT50 is widely considered to be the most prestigious ranking system for high quality academic business journals. The primary criteria for inclusion in the FT50 are generated almost exclusively from subjective human input via surveys of business school deans and editorial input by the Financial Times staff. While some more objective data may be considered *indirectly* in these rankings—e.g., number of citations, institutional prestige of authors, editorial board member affiliations and reputations, etc.—they are not included *directly* in any methodical, scientific manner. Basically, it is not an exaggeration to characterize the FT50 as a ranking generated by opinionated, collective subjective human judgment. The list of 50 journals vaunted by the FT50 is the byproduct of a relatively small number of raters weighing-in on what they consider to be quality journals. The standards of quality here are not explicit, transparent, or verifiable and have caused much heated debate (Anwar, 2021; Christensen Hughes, 2020; Jack, 2021; Linacre, 2021b; Rodenburg et al., 2021; Steingard & Linacre, in press). This paradigm shift is part of a larger "reorientation" in "the ecosystem of ranking, rating and accreditation institutions" comprising leadership education (Morsing, 2021: 10). Encouragingly, there is a clarion call for a sea change in moving from conventional notions of "quality" to "impact":

> The growing demand for societal impact of teaching, research, and operations necessitates fresh approaches to our analysis of business school rankings. I discuss the *Financial Times'* approach and the need for fresh methods, metrics, and standards (Jack, 2021).

The second dataset of 50 journals is hand-picked by the team working collaboratively at Cabells and SJU. Where the FT50 were chosen on subjective standards of quality, this impact set of 50 journals was chosen because of hypothesized alignment with the type of impact expressed in the SDGs. It is critical to note that both the FT50 and this handpicked impact list of journals are both determined subjectively and unmoored from any objective scientific process. This does not make these selections invalid or unjustified, but does remind us of their *human judgement origins.* They are both expert-system, human intelligence driven, rationalized by explicit and tacit decision-rules about what counts for quality (FT50) and SDG-related impact (Cabells-SJU 50). In effect, the human subjectivity of both the FT50 and SDG50 is evaluated objectively through the SDG Impact Intensity algorithm.

To test and gather our SDG Impact Intensity evaluations, we collected an average of 700 articles per journal for both datasets, all from a publishing date of 2015 to 2020, approximately 80,000 titles and abstracts totalling 13,600,000 words. For each journal, we collected title and abstract information.[3] There have been multiple strategies undertaken to collect this data. Where appropriate, we developed web scraping tools that automatically download the required content. Some of the web scrapers are designed specifically for the websites of the

[3] It is publishing data analytics industry practice to utilize titles, abstracts, and keywords as a composite data set for analyzing textually derived themes for academic journals in the aggregate and individual articles. Full-text searches of articles are impractical and inconsistent because of paywalls. In future iterations of the SDGIIM we intend to include article keywords to enrich our data set and fortify results. It is a promising development that several academic publishers are offering an SDG keyword selection function for submitting authors.

most popular publishers of the Cabells *Journalytics* (Cabells, webpage-b) where the 3,000 journals originated. These web scrapers allow us to gather data for a large portion of the set. Next, we leveraged CrossRef (Crossref, webpage), a database with records for over 18 million content items from 1,500 publishers and societies. Publishers supply CrossRef with article information for their journals, but supplying the abstracts is optional. We queried CrossRef using a journal's ISSN and retrieved its articles' titles and abstracts starting from 2015 where available.

We manually collected data for some of the more elusive journals in the Cabells 3,000 business and business-related journals set. We exported this data from the ABI/INFORM ProQuest Database (ProQuest, webpage) and from the Saint Joseph's University Post Learning Commons & Drexel Library Resources (webpage). ProQuest's ABI/INFORM Database is a popular resource for researchers to find scholarly journals, newspapers, reports, and datasets. Similar content and querying interfaces are available through most institutions.

4. Results: SDG Impact Intensity as a Defensible Construct

4.1 Criterion validity of SDG Impact Intensity

To assess SDGIIM in academic journals we begin with a simple hypothesis. Based on the topics, theories, methodologies, author(s) intent, and overall content of an academic journal, it can be evaluated for its degree of SDG intensity. Since academic journals, and their subcomponent articles are textual products, this assessment must have the capacity to examine a journal's composite text (article titles and abstracts) and extract the degree to which that journal is meaningfully aligned with the SDGs. It is fortunate that the SDGs themselves are very well articulated across 17 goals, 169 targets, and 231 indicators (for a consolidated integration of goals, targets, and indicators see The Danish Institute for Human Rights, webpage). The SDGs are very rich in linguistic expression and provide a robust textual dataset to analyze the journals.

Now we can more concretely articulate our experimental design to assess the criterion validity (American Psychological Association, webpage) of SDG Impact Intensity—how effectively can it predict the SDG Impact Intensity of a given set of journals? We hypothesize that the Cabells-SJU 50 will score higher (see Figure 4). If this hypothesis is proven true to some meaningful degree, it would demonstrate that the algorithm powering the SDGIIM validated the human judgement of what constitutes a journal that is SDG-intense. This would help establish a form of “ground truth” and explainability criteria (Ding et al., 2021; Guidotti, 2021; Yalcin et al., 2021) for evaluating the accuracy of rating algorithms and implementing the idea of criterion validity.

In the output produced by the algorithm, the overall SDG Impact Intensity rating captures the holistic impact of the journal across all 17 SDGs. Additionally, extra weight is given to the occurrences of the keywords that explicitly mention the SDGs, e.g., “SDGs,” “Sustainable Development Goals.” “Agenda 2030,” and others (keyword bank SDG_0). The argument here is that any journals that refer specifically to the SDGs as a holistic system of policies and practices merit more influence on results. While a journal dedicated to, for example, gender equality in management, would score highly, it would score even higher if referring directly to SDG #5: Gender Equality. The other three categories output the top 3 SDGs covered by the journal. This is useful to pinpoint more directly how a journal might specialize in contributing to the impact of particular SDGs; because the scope of the SDGs is so vast, no individual journal can be expected to address more than perhaps 2 or three SDGs with any depth of impact.

Of course, SDG Impact Intensity is more nuanced than just a one-to-one word correspondence and frequency count; a journal is more SDG-intense if there are simply more matching words in the overlapping set. Indeed, at a basic level, matching words and the quantity of those matching words provides a preliminary assessment of SDG-intensity. Yet, since SDG Impact Intensity is intended to be a metric of impact for journals vis-à-vis the SDGs, does a large overlap of matching words incontrovertibly correlate with increased SDG-intensity? There are many problems with this superficial approach (Jack, 2020). First, each of the SDG keywords is not equivalent in terms of their contribution to impact. Clearly, there are variations of meaning weights between words in the SDG keyword banks. And, especially for multiple word sequences in the keyword banks, journals may not exactly employ those terms, but still reflect the meaning of them as syntax matters. Negation, the notion that a keyword identified in a journal may be using the SDG keyword bank term in a contrary manner resulting in negative impact is a critical consideration. Also, there are simple contextual meaning attributions that can be wrong; e.g., “equity” can denote equitable distribution or financial capital, depending upon the context. We elaborate further on this topic in the final section.

4.2 Predictive validity of SDG Impact Intensity

As seen in Figures 5.1 and 5.2 below, our overall prediction was supported. Green journals are Cabells-SJU 50 selections and pink journals are the FT50. Only 2 (4%) of the FT50 journals scored in the top 50 of SDG Impact Intensity while 48 (96%) of the Cabells-SJU 50 scored in the top 50 of SDG Impact Intensity. Moreover, these 2 journals placed very low at positions 42 and 50. Only 4 (8%) of Cabells-SJU journal selections placed in the bottom 50 and none below position 64. This overwhelmingly demonstrates that, at least in discriminating between these two data sets, the decision-rule logic inherent in the SSDGIIM's algorithm effectively distinguishes journals chosen for SDG Impact Intensity. Since only 2 of the FT50 journals registered in the upper half of SDG Impact Intensity, we conclude that the decision-rules for choosing those 50 journals based on "quality" standards do not intentionally consider the "impact" considerations of SDG Impact Intensity. Of course, this conclusion does not advocate a wholesale discount of FT50 journals from aligning with the SDGs; evidentiarily, there are FT50 journals that do score at least some degree of SDG Impact Intensity, particularly around the lower midpoint of Figure 5.1. Yet, it is promising that the approach produced such a clear distinction, a foundation to feel confidence that the underlying logic of the SDG Impact Intensity is solid and can be developed.

		SDG Impact Intensity Wheel Rating	Top #1 SDG	Top #2 SDG	Top #3 SDG	Journal Type
1	Natural Resources Forum	3	11	12	7	SDG Impact Intensity
2	Environment, Development and Sustainability	3	11	12	6	SDG Impact Intensity
3	Globalization and Health	3	3	10	9	SDG Impact Intensity
4	Smart and Sustainable Built Environment	3	11	12	7	SDG Impact Intensity
5	International Journal of Sustainable Development and World Ecolog	3	11	12	1	SDG Impact Intensity
6	Gender in Management	3	10	5	8	SDG Impact Intensity
7	International Journal of Environmental Sustainability, The	2.5	11	12	6	SDG Impact Intensity
8	International Journal of Sustainability Policy and Practice, The	2.5	11	12	6	SDG Impact Intensity
9	International Journal of Environment & Sustainable Development	2.5	12	11	7	SDG Impact Intensity
10	Agronomy for Sustainable Development	2.5	12	15	2	SDG Impact Intensity
11	International Journal of Water Resources Development	2.5	6	12	11	SDG Impact Intensity
12	International Journal of Agricultural Sustainability	2.5	2	12	11	SDG Impact Intensity
13	Current Opinion in Environmental Sustainability	2.5	11	13	12	SDG Impact Intensity
14	Sustainability Science	2.5	11	12	13	SDG Impact Intensity
15	Journal of Hunger & Environmental Nutrition	2.5	2	12	10	SDG Impact Intensity
16	Sustainable Development	2.5	12	11	1	SDG Impact Intensity
17	Journal of Sustainability Education, The	2.5	4	16	10	SDG Impact Intensity
18	Sustainability	2.5	11	12	9	SDG Impact Intensity
19	Food Policy	2.5	2	12	10	SDG Impact Intensity
20	Hydrology and Earth System Sciences	2.5	6	11	12	SDG Impact Intensity
21	Journal of Renewable Energy	2.5	7	12	9	SDG Impact Intensity
22	Sustainability: Science, Practice, & Policy	2.5	12	11	9	SDG Impact Intensity
23	Journal of Sustainable Tourism	2.5	11	12	1	SDG Impact Intensity
24	World Journal of Entrepreneurship, Management and Sust. Dev.	2.5	8	9	10	SDG Impact Intensity
25	International Journal of Sustainable Energy	2.5	7	12	9	SDG Impact Intensity
26	Natural Resources Journal	2.5	12	11	7	SDG Impact Intensity
27	Nature Sustainability	2.5	12	11	13	SDG Impact Intensity
28	Sustainability Accounting, Management and Policy Journal	2.5	9	12	11	SDG Impact Intensity
29	International Journal of Social Sustainability in...	2.5	11	12	10	SDG Impact Intensity
30	Agroecology and Sustainable Food Systems	2.5	12	2	11	SDG Impact Intensity
31	Landscape and Urban Planning	2.5	11	6	15	SDG Impact Intensity
32	African Journal of Economic and Sustainable Development	2.5	8	1	10	SDG Impact Intensity
33	Journal of Sustainable Finance & Investment	2	7	13	12	SDG Impact Intensity
34	Business, Strategy and the Environment	2	12	9	7	SDG Impact Intensity
35	Journal of Agricultural and Environmental Ethics	2	12	12	15	SDG Impact Intensity
36	Sustainability: The Journal of Record	2	11	12	13	SDG Impact Intensity
37	Environmental Progress & Sustainable Energy	2	7	12	13	SDG Impact Intensity
38	International Food & Agribusiness Management Review	2	12	2	9	SDG Impact Intensity
39	Environment: Science and Policy for Sustainable Development	1.5	13	11	7	SDG Impact Intensity
40	Journal of Environmental Law	1.5	13	11	7	SDG Impact Intensity
41	Organization & Environment	1.5	12	10	11	SDG Impact Intensity
42	Research Policy	1.5	9	8	4	FT50
43	Process Safety and Environmental Protection	1.5	12	6	7	SDG Impact Intensity
44	Ethics, Policy & Environment	1.5	13	11	7	SDG Impact Intensity
45	Corporate Governance: The International Journal of Business Socie	1.5	16	9	10	SDG Impact Intensity
46	Corporate Governance	1.5	16	10	9	SDG Impact Intensity
47	Journal of Strategic Innovation and Sustainability	1.5	9	12	7	SDG Impact Intensity
48	Business & Society	1	10	9	8	SDG Impact Intensity
49	Social and Environmental Accountability Journal	1	13	9	16	SDG Impact Intensity
50	Management Information Systems Quarterly	1	9	4	10	FT50

Fig. 5.1 Prediction table for FT50 and Cabells-SJU 50 journals

		SDG Impact Intensity Wheel Rating	Top #1 SDG	Top #2 SDG	Top #3 SDG	Journal Type
51	Quarterly Journal of Economics	1	8	10	5	FT50
52	Law & Policy	1	16	10	5	SDG Impact Intensity
53	Journal of Business Ethics	1	10	9	8	FT50
54	Human Relations	1	8	10	5	FT50
55	Environmental Ethics	1	13	11	15	SDG Impact Intensity
56	Administrative Science Quarterly	1	8	10	9	FT50
57	Review of Finance	1	10	16	4	FT50
58	Business Ethics, the Environment, and Responsibility	1	10	16	9	SDG Impact Intensity
59	Information Systems Research	0.5	9	8	5	FT50
60	Manufacturing & Service Operations Management	0.5	12	9	8	FT50
61	Journal of International Business Studies	0.5	10	9	8	FT50
62	Strategic Entrepreneurship Journal	0.5	8	9	10	FT50
63	Journal of Applied Psychology	0.5	8	9	10	FT50
64	Journal of Leadership, Accountability and Ethics	0.5	10	16	4	SDG Impact Intensity
65	Review of Economic Studies, The	0.5	8	10	11	FT50
66	Organization Science	0.5	9	8	10	FT50
67	Journal of Management	0.5	9	8	5	FT50
68	Human Resource Management	0.5	8	10	9	FT50
69	Journal of Business Venturing	0.5	8	9	17	FT50
70	Journal of Political Economy	0.5	8	10	4	FT50
71	American Economic Review	0.5	8	10	5	FT50
72	Entrepreneurship Theory and Practice	0.5	8	10	9	FT50
73	MIT Sloan Management Review	0.5	8	10	9	FT50
74	Academy of Management Journal	0.5	9	8	10	FT50
75	Accounting, Organizations and Society	0.5	10	8	9	FT50
76	Journal of Operations Management	0.5	9	12	8	FT50
77	Production and Operations Management	0.5	12	9	8	FT50
78	Journal of Management Studies	0.5	9	8	10	FT50
79	Journal of Management Information Systems	0.5	9	8	10	FT50
80	Organization Studies	0.5	8	9	16	FT50
81	Journal of Consumer Research	0.5	9	12	8	FT50
82	Strategic Management Journal	0.5	9	8	10	FT50
83	Harvard Business Review	0.5	10	8	9	FT50
84	Journal of the Academy of Marketing Science	0.5	9	8	10	FT50
85	Journal of Consumer Psychology	0.5	9	12	8	FT50
86	Econometrica	0	8	10	1	FT50
87	Journal of Marketing	0	9	8	10	FT50
88	Journal of Marketing Research	0	9	12	8	FT50
89	Management Science	0	10	8	9	FT50
90	Organizational Behavior and Human Decision Processes	0	9	8	10	FT50
91	Journal of Finance	0	8	10	16	FT50
92	Academy of Management Review	0	9	8	10	FT50
93	The Journal of Financial Economics	0	16	10	8	FT50
94	Contemporary Accounting Research	0	9	16	8	FT50
95	Review of Accounting Studies	0	1	10	16	FT50
96	Marketing Science	0	12	9	10	FT50
97	Journal of Financial & Quantitative Analysis	0	10	16	8	FT50
98	Accounting Review, The	0	16	1	9	FT50
99	Journal of Accounting and Economics	0	8	9	10	FT50
100	Operations Research	0	12	1	9	FT50

Fig. 5.2 Prediction table for FT50 and Cabells-SJU 50 journals

The final SDG Impact Intensity ratings produced from column B of Figures 5.1 and 5.2 will be translated from their numerical basis into a six-element SDG wheel rating system as seen in Figure 6. SDG Impact Intensity represents a graphical adaptation of the number produced by the process described above. The measure is represented by a 3 SDG wheel rating system that subdivides into half wheels, ranging from one-half a wheel to three full wheels. The number and proportion of wheels assigned to the ratings is determined by the score produced by the algorithm. Cutoffs for the wheel ratings are based on an imposed curve for these scores. This visualization is akin to similar star rating systems found on e-commerce websites, movie reviews, restaurant reviews, etc. The introduction of the coarser 3 wheel rating system to represent the more finely grained score of a journal is motivated by our recognition that evaluations of SDG impact are inherently approximated regardless of the level of sophistication of the algorithms leveraged by a given instantiation of our framework. In fact, even human experts might disagree among themselves on the specific score assigned to a journal, but they are likely to agree in terms of the measure such as the 3 wheel rating system.

Fig. 6 SDG Impact Intensity wheel rating system

At the time of writing, these 100 ratings are in the process of being attached to individual records in Cabells' *Journalytics* (Cabells, webpage-b) and currently available on the SDG Impact Intensity Journal Ratings (webpage). These ratings, along with the three individual top SDG ratings, will assist users of Cabells to better ascertain the SDG Impact Intensity of academic journals. We look forward to feedback from the Cabells user community and other networks as methods to further validate the SDGIIM with additional expert input to refine our algorithm.

5. Future research considerations

This section offers two broad types of considerations based on the constructs and application in this paper. First, we offer some considerations related to the overall domain of the SDGIE framework and SDG evaluation as a vehicle for 'social good.' Second, we offer remarks on what we have learned (and still need to learn) about how our particular instantiation of the SDG Impact Intensity Model (SDGIIM), as one of many possible models in the SDGIE framework, can contribute to more theoretical rigor and predictive validity when evaluating SDG impact on any textual dataset employing an algorithmic technique.

It is important to note that, while not implemented in the current instantiation, the framework we propose allows for incorporating a machine learning component that can help identify particular disciplinary language domains in order to increase the robustness of the algorithm. In its current instantiation, the SDGIIM algorithm was developed as an implementation of decision-rules based on expert knowledge. This can be viewed as an instance of "learning by being told" (Mostow, 1983: 367) and was chosen in order to form a principled, transparent foundation. Given this solid foundation, more sophisticated AI and machine learning techniques can be integrated thanks to the flexibility of the SDGIIM. We provide here a brief discussion of some such techniques.

From a certain point of view, the algorithm's interpretation of the text in identifying the keywords is quite rudimentary, as it does not take into account the context of the sentences where the keywords are found. For example, "equity" can mean two very different things as illustrated earlier. Negative constructions may also need to be interpreted carefully. Extraneous words mixed within the words of a keyword phrase may also alter the meaning of the passage to the point that a human expert would not want to associate it with an occurrence of the keyword. Multiple techniques related to Natural Language Processing, including full-fledged Natural Language Understanding (Agarwal, 2019) and potentially Sentiment Analysis (Whitelaw et al., 2005), could be introduced to achieve a better evaluation of context and meaning.

The assignment of weights to keywords could also be refined by leveraging machine learning. Specifically, it would not be difficult to allow experts to provide their intended SDG wheel rating for a collection of journals and use machine learning techniques to adjust the weights accordingly. Conversational AI (Yousef & Torad, 2019) and Explainable AI (Adadi & Berrada, 2018) techniques could also be used to make the approach more transparent to users. For instance, a user may be surprised by the difference in ratings between two journals. A refinement of the current instantiation might allow the system to provide an explanation for the different ratings, for instance, by pointing out which keywords made a particular difference. The ensuing dialogue with the user might then allow the system to refine the evaluation strategy.

One overarching issue is framing the SDGIIM as a *standards-based system* of evaluation. The algorithm's standards are generated by the extant keyword bank and its further refinement as described above. The SDG Impact Intensity of a journal is effectively determined by the degree to which it reflects keywords—both in the number of keywords hit and the nature of those hits. With refining the standard by adding keywords through human input and automated machine learning, this poses a potential problem of relativism while considering the standard. The standard is changing to fit the admission of new journals and new keywords deemed to be aligned with the SDGs. At what point does this exceptionalism of human tweaking debase the objectivity of the algorithm? Of course, this iterative, evolutionary process is quite healthy. Any standard needs reassessment and tweaking over time as more is learned to inform it.

Yet, this begs the question of data comprehensiveness. How will we know when the SDGIIM has processed enough journals and attendant data to be considered a reliable technique? A practical application will illuminate this point. The next phase of research and deployment of the SSDGIIM will involve developing SDG Impact Intensity ratings for 8% (250/3000) academic business and business-related journals. Undoubtedly, the interrogation of this larger data set will most definitely result in the addition of keywords and phrases to the keyword bank and in the recalibration of the percentiles of ratings for specific journals. A journal currently in the upper quartile of our 100 journals may move down, considered amongst the 250. So, it appears that a journal's rating is relative to the overall data set and not anchored to a fixed standard. In fact, it is already the case that divisions of the SDG Impact Intensity wheel rating system are adjusted by human input. For instance, much thought was given to whether or not *any* journal should receive a 3 full wheel rating, implying that it was perfectly aligned with the SDGs. Would it be more beneficial to tweak the algorithm to prohibit any journals from achieving a 3 full wheel rating, suggesting that no journal ever fully supported the SDGs? Simply put, if you rate a fixed set of journals and you develop your standard from that ratings system, you are most likely to segregate that fixed set into comparison groups with some being better than others generating a relativistic evaluation. To combat this issue of shifting standards and perhaps arbitrarily established standards, the key will be to continuously develop the SDGIIM with more data comprehensiveness.

The above discussion of relativism and data comprehensiveness leads to another critical consideration of ratings vs. rankings. Currently, the SDGIIM is configured to provide a wheel rating of SDG Impact Intensity for the 100 journals in the data set. Each journal is evaluated *independently* through the keyword algorithm—an individual *rating.* However, as this project has unfolded, we have learned of a desire to convert the ratings produced by our system into a *relativistic ranking.* Let us take a moment to discern the crucial difference. A ranking would take a fixed set of journals—let us say the 100 in our project—utilize the algorithm to rank order the 100 from 1-100 with the #1 spot being occupied by the *most* SDG-intense journal and the 100th being the *least* SDG-intense. Such a ranking would provide a false sense of efficacy because the ranking, although based on the keyword bank and how the SDGIIM processes it, would skew toward forced ordination instead of allowing each journal's rating to stand on its own.

Moreover, if the ranking paradigm is applied to a particular set of journals (e.g., like the FT50) there will by definition be the #1 (best) SDG-intense journal and the #50 (worst) SDG-intense journal regardless of the objectively determined rating produced by the algorithm. By way of analogy, if one is asked to rank order the quality of 10 types of out-of-date, spoiled cheese, one will produce a list from 1-10, but #1 on the list will still be cheese unfit for human consumption. The point here is to ensure that future iteration of the SDGIIM and its algorithm focus on a ratings-based system rather than a rankings-based system.

As the SDGIE framework is a contribution to AI4SG (Floridi et al., 2020), its potential to galvanize a number of evaluative approaches involving SDGs is promising. As SDGs continue to gain importance in academia, business, government, etc., the SDGIE framework allows for myriad approaches to coalesce around key questions of constructs, methods, and AI techniques. Since the scope of this paper focuses on evaluating SDG Impact Intensity with our SDGIIM in the domain of academic journal publications, work to adapt other models and applications in the SDGIE framework could prove fruitful. We hope that our efforts in this paper can make a contribution to the evolution of how AI is applied to advancing the "'social good'" as detailed in the call for papers in this special issue (AI & SOCIETY, 2021: 1) offered by the SDGs in academic journal publishing and research, as well as other domains of application.

At the completion of Phase 1 that included generating predictive ratings for the first 100 journals as a sample, in a collaborative effort with Cabells, the Saint Joseph's University team formulated a 75/75 predictive test for Phase 2 to further refine the algorithm using machine learning models and further test the existing SDGIIM. Cabells provided the Saint Joseph's University team with a list of 15,414 business journals and SJU selected 150 with the following conditions. To ensure the prospective variance in sample selection (Bhandari, 2020) we developed a 75/75 predictive approach. For the first sample, we selected 75 journals (from a subset of business ethics, sustainability, and law journals) that the Saint Joseph's University team, based on our experience with the first 100 journals in Phase 1, predicted would score highly with SDG Impact Intensity. And for the second 75, we selected a mix of accounting journals expecting them to have lower SDG Impact Intensity ratings. Effectively, the Saint

Joseph's University team used its accumulated knowledge as subject matter experts to test the SDGIIM. Preliminary results from the 75/75 predictive test confirm the validity of the SDGIIM. As discussed above, special emphasis is given to inducing sample variation as it will allow us to develop better predictive models for SDG Impact Intensity ratings. As a future research consideration, we may conduct research on predicting journal ratings using the existing results as a base for machine learning models (Song et al., 2017).

In this paper, we examined how published journal research can be evaluated with automated and AI-based techniques, in terms of its contributions to aligning with and advancing the SDGs. We proposed the SDG-Intense Evaluation (SDGIE) framework as an organizing schema encompassing the current approaches to evaluating published journal research terms of their SDG-related contributions. Additionally, we described one particular instantiation of the framework, the SDG Impact Intensity model or SDGIIM, and applied it to the particular domain of academic journal ratings as a case study. We trust our paper adds insight to these exciting movements of AI4SG (Floridi et al., 2020) and SDGs and AI (Cowls et al., 2021) in order to tackle some of the world's most exigent human and environmental problems.

6. References

AAAI. (2017). *AI for Social Good*. AAAI 2017 Spring Symposia Registration. Retrieved July 29, 2021, from https://aaai.org/Symposia/Spring/sss17symposia.php#ss01

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/access.2018.2870052

Agarwal, M. (2019). An Overview of Natural Language Processing. *International Journal for Research in Applied Science and Engineering Technology*, *7*(5), 2811–2813. https://doi.org/10.22214/ijraset.2019.5462

AI & SOCIETY. (2021). Special issue on AI for People (with sponsored Best Paper Award). Springer. https://www.springer.com/journal/146/updates/18583616?error=cookies_not_supported&error=cookies_not_supported&code=b97d39db-75a5-4f36-aa87-1cf3f6cdae1e&code=44415b01-7edf-4834-9a3b-3ef34bbf2ed1

AISOC. (2017). *AI for Social Good (AISOC)*. University of Southern California. http://scf.usc.edu/%7Eamulyaya/AISOC17/index.html

American Psychological Association. (webpage) (2020). *APA Dictionary of Psychology - Criterion validity*. Retrieved July 29, 2021, from https://dictionary.apa.org/criterion-validity

Anwar, C. M. (2021, January 1). *Shifting Paradigm from For-Profit Journal Indexing to Not-for-Profit Academic Journal Quality/Ranking Lists.* https://www.academia.edu/50244184/Shifting_Paradigm_from_For_Profit_Journal_Indexing_to_Not_for_Profit_Academic_Journal_Quality_Ranking_Lists?email_work_card=reading-history

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. SAGE Open, 9(1). https://doi.org/10.1177/2158244019829575

Bhandari, P. (2020, October 12). *Understanding and calculating variance* [online forum post]. https://www.scribbr.com/statistics/variance/

Bhargava, V. R., & Velasquez, M. (2021). Ethics of the attention economy: The problem of social media addiction. *Business Ethics Quarterly*, *31*(3), 321-359.

Bornmann, L., Haunschild, R., & Adams, J. (2019). Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF). *Journal of Informetrics*, *13*(1), 325–340. https://doi.org/10.1016/j.joi.2019.01.008

Cabells. (webpage-a). (2021) *About us*. Retrieved July 29, 2021, from http://www2.cabells.com/about

Cabells. (webpage-b). (2021) *Journalytics*. Retrieved July 30, 2021, from https://www2.cabells.com/about-journalytics

Christensen Hughes, J. (2020, April). *The Future of Business Schools Rankings & Ratings - Davos 2020 Report*. Oikos. https://issuu.com/oikos-world/docs/davosrecapreport2020

Chui, M., Harrysson, M., Manyika, J., Roberts, R., Chung, R., Nel, P., & van Heteren, A. (2019, November 20). *Applying artificial intelligence for social good*. McKinsey & Company. https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good#

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence*, *3*(2), 111–115. https://doi.org/10.1038/s42256-021-00296-0

Crossref. (2021) *You are Crossref - Crossref*. Retrieved July 30, 2021, from https://www.crossref.org

Digital Science & Research Solutions. (2021) *Dimensions AI*. Dimensions AI. Retrieved July 29, 2021, from https://app.dimensions.ai/discover/publication

di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, *121*(December 2020), 283–314. https://doi.org/10.1016/j.jbusres.2020.08.019

Ding, Y., Botzer, N., & Weninger, T. (2021). Posthoc Verification and the Fallibility of the Ground Truth. ArXivLabs. Published. https://arxiv.org/abs/2106.07353v1

Elsevier. (2021). SDG Research Mapping Initiative. Elsevier.Com. Retrieved July 30, 2021, from https://www.elsevier.com/about/partnerships/sdg-research-mapping-initiative

Emerald Insights. (2021). *Concise Guides to the United Nations Sustainable Development Goals*. Retrieved July 29, 2021, from https://www.emerald.com/insight/publication/acronym/SDG

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, *26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Garwood, K., Steingard, D., & Balduccini, M. (2020). *Dynamic Collaborative Visualization of the United Nations Sustainable Development Goals (SDGs): Creating an SDG Dashboard for Reporting and Best Practice Sharing.* SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0009172302940300

Gomes, C., Dietterich, T., Barrett, C., Conrad, J., Dilkina, B., Ermon, S., Fang, F., Farnsworth, A., Fern, A., Fern, X., Fink, D., Fisher, D., Flecker, A., Freund, D., Fuller, A., Gregoire, J., Hopcroft, J., Kelling, S., Kolter, Z., . . . Zeeman, M. L. (2019). Computational sustainability: Computing for a Better World and a Sustainable Future. *Communications of the ACM*, *62*(9), 56–65. https://doi.org/10.1145/3339399

Google. (2021). *Top Publications*. Google Scholar. https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=en

Google AI. (2021). *AI for Social Good - Applying AI to some of the world's biggest challenges*. Retrieved July 29, 2021, from https://ai.google/social-good/

Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*, *18*(1). https://doi.org/10.1016/j.ijme.2019.100330

Guidotti, R. (2021). Evaluating local explanation methods on ground truth. *Artificial Intelligence,* 291. https://doi.org/10.1016/j.artint.2020.103428

Hager, G., Drobnis, A., Fang, F., Ghani, R., Greenwald, A., Lyons, T., C. Parkes, D., Schultz, J., Saria, S., F. Smith, S., & Tambe, M. (2017, March). *Artificial Intelligence for Social Good* (Grant No. 1136993). ArXiv.org. https://arxiv.org/ftp/arxiv/papers/1901/1901.05406.pdf

Harley, B., & Fleming, P. (2021). Not Even Trying to Change the World: Why Do Elite Management Journals Ignore the Major Problems Facing Humanity? *The Journal of*

Applied Behavioral Science, *57*(2), 133–152. https://doi.org/10.1177/0021886321997189

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January). Word Cloud Explorer: Text Analytics Based on Word Clouds. *2014 47th Hawaii International Conference on System Sciences*, 1833–1842. https://doi.org/10.1109/hicss.2014.231

Jack, A. (2020, December 6). Weighing up business schools' work on sustainability. *Financial Times.* https://www.ft.com/content/6b499b5b-76fc-4fee-9684-f8055e52c46e

Jack, A. (2021). Business School Rankings: The Financial Times' Experience and Evolutions. *Business & Society*. https://doi.org/10.1177/00076503211016783

Kriebitz, A., & Lütge, C. (2020). Artificial Intelligence and Human Rights: A Business Ethical Assessment. *Business and Human Rights Journal*, *5*(1), 84–104. https://doi.org/10.1017/bhj.2019.28

Kulevicz, R. A., Porfirio, G. E. D. O., de Oliveira, O. S., Zavala Zavala, A. A., Silva, B. A. D., & Constantino, M. (2020). Influence of sustainability reports on social and environmental issues: bibliometric analysis and the word cloud approach. Environmental Reviews, 28(4), 380–386. https://doi.org/10.1139/er-2019-0075

Linacre, S. (2021a, March 17). *The Source/Cabells launches new SDG Impact Intensity* journal rating system in partnership with Saint Joseph's University's Haub School of Business. Cabells.

Linacre, S. (2021b, March 31). *The Source/Opening up the SDGs*. The Source. https://blog.cabells.com/2021/03/31/opening-up-the-sdgs/

Linacre, S. (2021c, April 28). *The Source/What really counts for rankings?* The Source. https://blog.cabells.com/2021/04/28/what-really-counts-for-rankings/

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval* (1st Edition). Cambridge University Press. https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html

Morsing, M. (2021). PRME–principles for responsible management education: Towards transforming leadership education. In *Responsible Management Education* (pp. 3-12). Routledge.

Mostow, D. J. (1983). Machine Transformation of Advice Into a Heuristic Search Procedure. *Machine Learning*, *In: Michalski R.S., Carbonell J.G., Mitchell T.M. (eds) Machine Learning. Symbolic Computation. Springer, Berlin, Heidelberg*, 367–403. https://doi.org/10.1007/978-3-662-12405-5_12

Nakamura, M., Pendlebury, D., Schnell, J., Szomszor, M., ISI Institute for Scientific Information, & Web of Science Group. (2019, April). *Navigating the Structure of Research on Sustainable Development Goals*. Clarivate. https://clarivate.com/webofsciencegroup/campaigns/sustainable-development-goals/

Ormans, L. (2016, September 12). *50 Journals used in FT Research Rank*. Financial Times. https://www.ft.com/content/3405a512-5cbb-11e1-8f1f-00144feabdc0

Pontille, D., & Torny, D. (2010). The controversial policies of journal ratings: evaluating social sciences and humanities. Research Evaluation, 19(5), 347–360. https://doi.org/10.3152/095820210x12809191250889

Post Learning Commons & Drexel Library. (webpage). *Post Learning Commons & Drexel Library | Saint Joseph's University*. https://sites.sju.edu/library/

ProQuest. (2021). *LibGuides: ABI/INFORM Collection: Home*. Retrieved July 30, 2021, from https://proquest.libguides.com/abiinformcollection

Purdy, M. (2020). Unlocking AI's Potential for Social Good. *Harvard Business Review*, *2020*(October 27). https://hbr.org/2020/10/unlocking-ais-potential-for-social-good

Purnell, P. J. (2022). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case Study of SDG 13: Climate Action. *arXiv preprint arXiv:2201.02006*.

Ravenscroft, J., Liakata, M., Clare, A., & Duma, D. (2017). Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. *PLOS ONE*, *12*(3), e0173152. https://doi.org/10.1371/journal.pone.0173152

REF. (2020). *REF 2021*. Retrieved July 30, 2021, from https://www.ref.ac.uk

RELX. (2022). *SDG Resource Centre - Leading-edge information on the Sustainable Development Goals*. Retrieved January 24, 2022, from https://sdgresources.relx.com/

Responsible Research in Business Management [RRBM]. (2021). A vision for responsible research in business management. RRBM Network. Retrieved July 30, 2021, from https://www.rrbm.network/

Rodenburg, K., De Silva, V., & Christensen Hughes, J. (2021). SDGs: A Responsible Research Assessment Tool toward Impactful Business Research. *Sustainability*, *13*(24), 14019.

Rotterdam School of Management. (2021). *SDG ranking*. Retrieved July 29, 2021, from https://rsmmetrics.nl/sustainable-development-goals/triple-crown-sdg/journals-3

Schemm, Y. (2020, September). *Report: Mapping research to advance the SDGs*. Elsevier. https://www.elsevier.com/connect/sdg-report

ScienceOpen. (2021). *UN Sustainable Development Goals on ScienceOpen*. Retrieved July 29, 2021, from https://www.scienceopen.com/collection/e67a99d4-ef59-42f8-a498-18ec810fd9ac

SDSN Australia, New Zealand & Pacific. (2021, March 5). *Resources*. https://ap-unsdsn.org/resources/

Song, C., Ristenpart, T., & Shmatikov, V. (2017). *Machine Learning Models that Remember Too Much: Vol. CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (S. Congzheng, T. Ristenpart, & V. Shmatikov, Eds.; Issue October 2017). ACM. https://doi.org/10.1145/3133956.3134077

Springer Nature. (2021). *The Sustainable Development Goals Programme*. Retrieved July 29, 2021, from https://www.springernature.com/gp/researchers/sdg-programme

Steingard, D. & Linacre, S. (in press, 2022). Transforming academic journal assessment from "quality" to "impact": A case study of the SDG Impact Intensity academic journal rating artificial intelligence system. In C. Hauser and W. Amann (Eds.), In *Responsible management education and the digital transformation challenge.* Palgrave Macmillan.

Taylor & Francis. (2022). *Sustainable Development Goals online.* Informa. Retrieved January 24, 2022, from https://app.dimensions.ai/discover/publication

The Danish Institute for Human Rights. (webpage). *Goals, targets and indicators|The Human Rights Guide to the Sustainable Development Goals*. Retrieved July 29, 2021, from https://sdg.humanrights.dk/en/goals-and-targets

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., C.M. Belgrave, D., Ezer, D., Cachat Van Der Haert, F., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurina, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., de Wever, W., Clopath, C. et al. (2020, August 7). AI for social good. *Nature Communications*, *11*. https://www.nature.com/articles/s41467-020-15871-z?error=cookies_not_supported&code=0de573ec-82a3-404e-8cdb-d08423d4c023

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021, February 20). The ethics of algorithms: key problems and solutions. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01154-8

United Nations. (2020). *SDG Publishers Compact*. United Nations Sustainable Development. Retrieved July 29, 2021, fromhttps://www.un.org/sustainabledevelopment/sdg-publishers-compact/

United Nations. (2015). *THE 17 GOALS - Sustainable Development*. Retrieved July 29, 2021, from https://sdgs.un.org/goals

United Nations. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. Retrieved July 29, 2021, from https://sdgs.un.org/2030agenda

University of Worcester & Kingston University London. (webpage). *Keywords*. Harrison K Saperstein. Retrieved July 29, 2021, from https://hksaperstein.github.io/The-Big-Benchmarking-Tool/html/keywords.html

Véliz, C. (2021). Moral zombies: why algorithms are not moral agents. *AI & SOCIETY*, *36*(487–497). https://doi.org/10.1007/s00146-021-01189-x

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-019-14108-y

Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management - CIKM '05, October 2005*, 625–631. https://doi.org/10.1145/1099554.1099714

Wylie, C., Zuboff, S., Kaiser, B., Zuboff, T. A. O. S. C. B. S., & 978–1541758001. (2022). *Mindf*ck, The Age of Surveillance Capitalism, Targeted [Hardcover] 3 Books Collection Set*. Profile Books/HarperCollins.

Yalcin, O., Fan, X., & Liu, S. (2021). Evaluating the Correctness of Explainable AI Algorithms for Classification. ArXivLabs. https://arxiv.org/abs/2105.09740

Yousef, M., & Torad, M. A. (2019, November). A Treatise On Conversational AI Agents: Learning From Humans' Behaviour As A Design Outlook. *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, United Arab Emirates. https://doi.org/10.1109/icecta48151.2019.8959585