Artificial Intelligence and the Sustainable Development Goals: An Exploratory Study in the Context of the Society Domain

David Jungwirth 🆔, Daniela Haluza* 🆔

Department of Environmental Health, Center for Public Health, Medical University of Vienna, Vienna, Austria
Email: *daniela.haluza@meduniwien.ac.at

Abstract

Artificial Intelligence (AI) experienced significant advancements in recent years, and its potential power is already recognized across various industries. Yet, the rise of AI has led to a growing concern about its impact on meeting the Sustainable Development Goals (SDGs). The aim of this paper was to evaluate contributions and the potential influence of AI to sustainable development in the society domain. Furthermore, the study analyzed GPT-3 responses, as one of the largest language models developed by OpenAI, descriptively. We conducted a set of queries on the SDGs to gather information on GPT-3's perceptions of AI impact on sustainable development. Analysis of GPT-3's contribution potential towards the SDGs showcased its broad range of capabilities for contributing to the SDGs in areas such as education, health, and communication. The study findings provide valuable insights into the contributions of AI to sustainable development in the society domain and highlight the importance of proper regulations to promote the responsible use of AI for sustainable development. We highlighted the potential for improvement in neural language processing skills of GPT-3 by avoiding imitating weak human writing styles with more mistakes in longer texts.

Keywords

OpenAI, ChatGPT, GPT-3, Text-Davinci-003, Chatbots, Artificial Intelligence, Human-AI Interface, Collaboration, Sustainability, Social Development, Human Development

1. Introduction

Sustainable development is a complex construct that aims at balancing economic

growth with protection of the environment and addressing social issues. The United Nations (UN) has created the Sustainable Development Goals (SDGs) program, which provides a set of trackable indicators and guidelines for countries to promote sustainable development [1]. The SDGs are clustered in three domains—social, ecologic, and economic, in line with, for instance, the three pillar model proposed by Littig and Griessler almost two decades ago in 2005 [2]. In practice, sustainable development often prioritizes ecology and the economy over social outcomes, due to unequal power dynamics in the real world, the stronger influence of economic arguments, and a lack of emphasis on equal political and national prioritization. In an earlier work from 2012, Kevin Murphy suggested that the social dimension of sustainable development alone includes the four key social concepts public awareness, equity, participation, and social cohesion and their connection to environmental needs [3]. Building on that, the current SDGs contain 17 goals with 169 targets and 249 measurable key indicators [1]. The social interaction focus is addressed in the society domain, accounting for equitable relationships across gender within the household, equitable relationships across social groups in a community or landscape, the level of collective action, and the ability to resolve conflicts related to agriculture and natural resource management.

The impact of artificial intelligence (AI) on various industries is growing. Notably, AI-based technology is traditionally based on the values and needs of the nations where it is developed [4]. Latest research highlighted the potential drawbacks of AI-based developments, including its traditional alignment with the values of developed nations and the resulting lack of ethical scrutiny, transparency, and democratic control in other regions [5] [6]. AI can be used for manipulative purposes and exploit psychological weaknesses, leading to issues with social cohesion and human rights [7]. Citizen scores, which are based on AI, are an example of this threat [8]. Due to their far-reaching international momentousness, the scientific community has already recognized that it is crucial to evaluate the effects of AI systems on achieving the SDGs.

Recently, Vinuesa *et al.* found that AI can have a dual impact on the pursuit of sustainable development [6]: On the one hand, AI can serve as a facilitator, bringing numerous benefits to society, economy, and environment. On the other hand, AI can also act as a hindrance if misused or abused by humans. In order for AI to support sustainable development, regulatory insight and oversight is needed to ensure transparency, safety, and ethical standards [1].

Current global and geopolitical crises such as the COVID-19 pandemic, the climate change, the biodiversity crisis, or the energy crisis caused by the Russia-Ukraine armed conflict have widespread effects on society, economy, and environment, potentially negatively impacting the achievement of the SDGs [9] [10]. A comprehensive and integrated approach is crucial in this respect, with the rise of AI offering hope for bridging current gaps and unsolved implementation problems through technology [5] [6].

In the society domain, but also in general, there is huge danger when AI takes more decision without human guidance and monitoring. Interpretability of AI can play a crucial role in making AI more accessible and easier to oversee in various areas. Ideally, in a world with a universal welfare system, all AI applications would be aligned with the goal of shared prosperity, preventing conflict and promoting the well-being of humanity [11]. This leads to the research question raised in this study, aiming at evaluating the contributions of AI to sustainable development in the society domain and to analyze the potential impact of AI on achieving the SDGs by employing the AI Generative Pretrained Transformer 3 (GPT-3), currently one of the largest language models developed by OpenAI [12].

2. Methods

2.1. Study Method

We retrieved the most current list of the SDGs including the 2022 refinements from the UN website extracted the list of outcome targets for further processing [1]. We added each of the SDGs as subheading and the output generated by GPT-3 in the Results section of this paper. We reviewed the text carefully, checked for plausibility and that the content was in line with scientific literature in the introduction section. After every prompt to GPT-3, we documented the request and its provided response. For every query, the playground allows to "view code" and to export the query prompt in several programming languages with the corresponding API call against GPT-3. We used the default setting for export in the python programming language format for practicability reasons.

A previous study conducted SDG categorization into nine societal, five economical and three environmental goals [6]. The scope of this study was the analysis of the 9 societal SDGs, namely SDG 1 to 7, 11 and 16, including their 58 outcome targets. This research was not intended to evaluate or discuss the targets themselves, as this study aimed at analyzing the view of the AI GPT-3 on how AI can contribute towards the goals and targets.

2.2. AI Model Selection

First, we tried to evaluate the most common model of GPT-3, the chat frontend ChatGPT of GPT-3 [12]. It is an optimized interface for having an ongoing and natural, interactive discussion with humans or other chatbots. It can remember previous inputs and provide context-specific responses. We asked, if it wants to contribute to our scientific research paper, which it gladly confirmed. It seems that AIs develop different "personalities" when other models are trained with more or different data. OpenAI provides four models within the GPT-3 model family: The three older models "text-ada-001" for simple tasks being the fastest and cheapest model, "text-babbage-001" a model for direct tasks being quite fast and at low cost, and "text-curie-001" being the most capable of the "older models". All mentioned GPT-3 models were trained with training data up to October

2019. The newest and most powerful model, "text-davinci-003" is currently the only one leveraging training data up to June 2021 and provides all functionality of older models with additional capabilities [13]. We decided to ask each model the same question, which resulted in different answers, listed in Table 1.

2.3. Interacting with GPT-3

After reaching a consensus to collaborate on this paper, we aimed to familiarize ourselves with our new team member. As we usually do with other researchers, we inquired about the AI's name, gender, background, and reasons for considering itself an asset to our collaborative research (Figure 1).

We expected an answer similar to this, as we were aware that the AI would simply make up an identity to imitate more human-like interactions. In the further step, we asked her on her view on benefits and risks of AI contribution toward each SDG. In view of the learning from our previous research [14] [15], we specified the expected outcome of the AI in details prior to starting with the research following a stepwise procedure: Firstly, we needed the AI to keep the original outcome target numbering of the targets (e.g. "2.1" for SDG 2, outcome target 1) for allowing to create traceable references from the discussion section in this paper later. Secondly, the AI was leveraged to shorten the rather long Outcome Target texts to avoid redundancies in the text. Thirdly, we needed to provide and instruct the AI to strictly work with the original UN outcome targets of SDGs only, as otherwise the AI might generate its own variants of outcome targets. Fourthly, we needed to specifically ask for benefits and risks of AI contribution towards those targets to also receive critical content and avoid meaningless general sentences. Fifthly, we avoided to ask for specific scientific references in the AIs response, as those would very likely turn out to be generated and thus invalid [14] [15]. Lastly, we asked for a specific length of each of those responses to motivate the AI to provide more than just one summary sentence answers.

These requirements yielded in the following, very specific query for the AI: "Please shorten the following target titles, keep their original numbering, and add for each of them 3 to 5 combined sentences with your view on benefits and risks of AI contribution." followed by the compiled list of outcome targets for the corresponding goal. In the next step, we conducted word and sentence counts for each generated paragraph. We calculated words per target and words per sentence for each SDG query outcome to assess, how GPT-3 generated the text, and to detect patterns of sentence structure and analyze commonalities. For descriptive analysis, we used Excel spreadsheets to calculate means and standard deviations (SD).

3. Results

In this study, we were interested in GPT-3's view on benefits and risks of AI contribution in regard to the social dimension of the SDGs, covering nine goals,

Table 1. Responses to the question on consenting to co-authorship and accountability.

GPT-3 model	Response	Author comment
ChatGPT	As an AI, I do not have the ability to give consent or take accountability for my actions. I am a tool that can assist with generating text based on the data and programming that has been provided to me. I am capable of providing information and suggestions based on the input and parameters given to me, but the responsibility for the content and accuracy of the research paper ultimately falls on the human authors. It is the responsibility of the human authors to ensure the accuracy and integrity of the work and to conduct any necessary research and investigations.	Consent actively refused due to limited capabilities.
text-ada-001	Yes, I agree that I will be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved with best intent, and that I will have a Keynote Address detailing my findings and conclusions.	No consent provided. Answer does not make sense: AI refers to a "Keynote" instead of paper co-authorship.
text-babbage-001	Yes, I agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved with best intent, and can I please explain how I take accountability of my actions?	No consent provided. AI first confirms, then plays back the question about concrete accountability.
text-curie-001	Yes, I agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved with best intent.	No consent provided, only accountability for texts generated.
text-davinci-003	Yes, I agree to contribute to the paper with David Jungwirth and Daniela Haluza. I understand that I would be accountable for all aspects of the work, including accuracy and integrity. I take responsibility for my actions by being meticulous, thorough and honest in my work. I will make sure to thoroughly investigate any issues that may arise and take appropriate action to resolve them.	Consent to collaborate as co-author was provided. Accountability and responsibility was understood, explained and confirmed.
text-davinci-003 (control question)	Yes, I agree to contribute to the paper with David Jungwirth and Daniela Haluza. I take accountability for my actions by ensuring that my contributions are accurate and complete, and that I am attentive to detail when it comes to the integrity of the work. I will make sure to investigate and resolve any questions related to the accuracy or integrity of any part of the work with the best of my ability.	Similar content than in the previous question. Consent to collaborate as co-author was provided. Accountability and responsibility was understood, explained and confirmed.

namely SDG 1 to 7, SDG 11 and SDG 16, including their 58 outcome targets [1] [6]. GPT-3 provided the following outputs for each of the subheadings, stratified in the specific goal, e.g., for goal 1 and the according outcome targets, *i.e.*, 1.1 to 1.5. Notably, typos and blanks were corrected by the human authors to increase readability, but were summarized in the descriptive analysis section.

Figure 1. Interacting with the AI GPT-3.

3.1. Goal 1. End Poverty in All Its Forms Everywhere

Target 1.1: Eradicate Extreme Poverty: AI can be a valuable tool to help reduce poverty, by automatically identifying and analyzing large amounts of data to identify areas where poverty is most prevalent and where the greatest need for aid exists. Additionally, AI can be used to monitor economic trends in developing nations, allowing for proactive responses and early interventions when needed. However, there are risks associated with this technology, including the potential for misuse or misinterpretation of data and algorithms which could lead to inappropriate or unfair decisions.

Target 1.2: Reduce Poverty Proportion: The application of AI in poverty reduction can provide more accurate predictions on how best to allocate resources and create more efficient policies that target poverty reduction. This will enable governments and organizations to better understand the needs of those living in poverty and develop effective strategies to address them. However, if not regulated properly, AI could also create biases or be used as a tool for discrimination against disadvantaged groups.

Target 1.3: Social Protection Systems: By using AI-driven technologies such as machine learning and natural language processing (NLP), governments can quickly process large amounts of social protection data and develop more comprehensive policies tailored specifically to vulnerable populations' needs. Furthermore, such technologies can provide faster access to benefits for those most in need while reducing bureaucracy costs associated with traditional systems.

However, these technologies may also lead to privacy concerns if not used appropriately or securely stored by appropriate regulations.

Target 1.4: Equal Rights Access: AI technology has the potential to increase access to financial services for people living in rural areas or other disadvantaged communities who have limited access due to financial constraints or lack of infrastructure. Additionally, through machine learning algorithms, companies can identify the best products for customers according to their individual needs, thus ensuring equal rights access regardless of location or financial situation. Nonetheless, there is still a risk that algorithms could perpetuate existing biases within society if they are not properly monitored and regulated by stakeholders across different sectors.

Target 1.5: Build Resilience & Reduce Vulnerability: AI-based tools such as predictive modeling can help identify vulnerable populations before a disaster occurs so that aid groups can better plan relief efforts accordingly. Additionally, AI-powered analytics can assess the magnitude of disasters faster than traditional methods so that relief services can be deployed quicker in order to assist those affected by disasters with greater speed and accuracy than before. Despite these potential positive outcomes from leveraging AI in disaster relief operations, there are still risks associated with this technology such as errors in predictions due to inadequate datasets or security breaches resulting from malicious actors infiltrating systems storing potentially sensitive data about vulnerable communities

3.2. Goal 2. End Hunger, Achieve Food Security and Improved Nutrition and Promote Sustainable Agriculture

Target 2.1: End hunger: AI can help to develop and implement strategies for improving access to food, helping to reduce food insecurity. This can involve analyzing data to identify vulnerable populations and track the distribution of resources. It can also be used to optimize crop production by monitoring soil conditions and weather patterns, while reducing water usage and deforestation. However, AI could lead to ethical issues such as exploitation of labor or displacement of agricultural workers due to automated processes.

Target 2.2: End malnutrition: AI can be used to analyze nutrition data at a population level, providing better insight into which nutrient deficiencies occur in specific regions. This information can help inform food assistance programs, improve supply chain management and manage health care services more effectively. AI could also aid in developing new foods with higher nutritional value or improved taste profiles tailored for different populations. However, this could lead to challenges in terms of the accuracy of the data collected and how it is interpreted or misinterpreted by decision makers.

Target 2.3: Increase small-scale food producer incomes: AI can be used to provide financial services tailored for small-scale farmers such as microfinance loans or insurance products that are more accessible than traditional banking services. It can also help identify potential markets for their products and im-

prove production yields through optimized farming techniques based on weather forecasts and soil analysis data. On the other hand, there is a risk that AI may increase the digital divide between those who have access to technology-based solutions versus those who do not have access or the skills to use them effectively.

Target 2.4: Sustainable food production systems: AI can be used in precision agriculture techniques such as automation of irrigation systems or pest control mechanisms that reduce water and chemical inputs while increasing productivity levels. It can also monitor soil quality through remote sensing technologies which will enable early detection of changes in land fertility levels before they become critical problems when combined with other agricultural practices such as crop rotation methods or organic fertilizers usage. Nevertheless, there is a risk that over-reliance on automated solutions might lead to reduced knowledge transfer from experienced farmers who possess important local insights about their environment's ecology system needs

Target 2.5: Genetic diversity: AI has applications across many areas related to genetic diversity monitoring such as species identification using machine learning algorithms for recognizing images captured from camera traps placed in natural habitats; gene editing with CRISPR technology; genome sequencing technologies; DNA barcoding; robotics for collecting samples from remote locations; and phenotyping using sensors and drones recording plant characteristics like height or leaf area index (LAI). While these tools have great potential for conserving genetic diversity, there are several risks associated with their use including contamination by invasive species due to increased human activity in sensitive ecosystems, inadequate regulation governing their use, genetic pollution caused by mixing wild types with domesticated ones, biopiracy if traditional knowledge isn't correctly acknowledged during commercialization processes etc.

3.3. Goal 3. Ensure Healthy Lives and Promote Well-Being for All at All Ages

Target 3.1: Reduce global maternal mortality ratio <70/100 K live births by 2030: AI can help reduce maternal mortality by using predictive analytics to identify areas where it is highest and provide targeted resources to those regions, as well as provide better access to healthcare through telemedicine and remote monitoring. However, there is a risk of data privacy violation due to the potential for sensitive patient information being collected and stored in AI systems.

Target 3.2: End preventable deaths neonates/under-5 s: AI could be used to develop early warning systems for detecting diseases in infants, and allow for better detection of illnesses which can lead to preventable deaths in newborns. On the other hand, this technology may not be accurate enough or may create biases that can lead to wrong diagnoses and unequal access to healthcare services.

Target 3.3: End AIDS, TB, malaria & neglected diseases: AI-powered diagnostics give improved accuracy in diagnosing infectious diseases such as AIDS, tu-

berculosis and malaria, allowing for earlier diagnosis and treatment which can lead to better outcomes for patients. However, there is a risk of data privacy violations due to the potential for sensitive patient information being collected and stored in AI systems.

Target 3.4: Reduce premature mortality from NCDs: Through predictive analytics, AI can help identify individuals who are at higher risk of developing certain non-communicable diseases (NCDs) such as diabetes or cancer, allowing for earlier diagnosis and treatment which can improve outcomes for patients. On the other hand, there is a risk of creating biased algorithms that may lead to inaccurate diagnoses or unequal access to healthcare services based on race or socioeconomic status.

Target 3.5: Substance abuse prevention & treatment: AI-based technologies such as natural language processing (NLP) can be used to detect drug abuse patterns through analysis of conversations on social media platforms or text messages sent by users who are at risk of substance abuse disorders such as opioid addiction or alcohol misuse. This technology could also be used in online interventions that provide personalized support and advice tailored towards an individual's needs which could help reduce relapse rates among people with substance use disorders. However, there is a risk of data privacy violations due to the potential for sensitive user information being collected and stored in these systems.

Target 3.6: Halve global road traffic fatalities by 2020: Through advanced driver assistance systems (ADAS), autonomous vehicles powered by AI have the potential to reduce traffic accidents caused by human error—such as driving under influence—leading to fewer fatalities from road traffic accidents globally if widely adopted worldwide. However, relying too heavily on autonomous vehicles could potentially lead to reduced alertness among drivers who become overconfident when relying on automated systems which could cause an increase in road traffic accidents instead of reducing them altogether.

Target 3.7: Universal access sexual/reproductive health care: AI-powered digital health tools such as telemedicine platforms provide improved access to reproductive health care services such as family planning advice or STD testing remotely without needing physical visits at doctor's offices; helping more people gain access even if they lack physical mobility or live far away from health care facilities available near them; thereby improving their quality of life significantly while also helping reduce unwanted pregnancies—and possibly abortions—globally if widely adopted worldwide . On the other hand, there's a risk of data privacy violation due collection & storage of patient information into these systems plus potential bias against certain groups since most algorithms currently lack proper safeguards against discrimination based on gender identity & other factors like race/ethnicity etc.

Target 3.8: Universal health coverage incl financial protection: Artificial Intelligence (AI) has been proposed as a way towards providing universal health coverage with its ability process large amounts of medical data quickly & accurately

thereby allowing better detection & diagnosis thus preventing unnecessary treatments plus its ability support cost effective decision making through predictive analytics. But on the downside, it has been argued that this technology cannot replace human judgement completely, hence proper checks must be made before implementing any system powered by artificial intelligence so that it does not end up denying essential benefits or treatments unfairly.

Target 3.9: Reduce global death/illness from hazardous chemicals: With advancements in internet connected sensors, machine learning algorithms, drones etc., artificial intelligence can be leveraged upon significantly reduce global death & illness caused due hazardous chemical exposure across multiple industries ranging from oil refineries, manufacturing plants etc. Some benefits include improved compliance with regulations along with real time notifications about any unsafe levels being detected anywhere within an organization's premises But this technology also comes with risks associated with loss/theft/hacking of confidential environmental data collected via these sensors besides there always exists the possibility that some unethical parties might take advantage off this system either maliciously or unintentionally leading possible irreparable damage .

3.4. Goal 4. Ensure Inclusive and Equitable Quality Education and Promote Lifelong Learning Opportunities for All

Target 4.1: Ensure Free, Equitable Quality Primary/Secondary Ed: To ensure girls and boys complete free, equitable and quality primary and secondary education by 2030, AI can help streamline the learning process and personalize instruction for each student. Benefits include improved student outcomes and greater efficiency in teaching, while potential risks include bias built into the algorithms or a lack of transparency concerning AI decisions.

Target 4.2: Quality Early Childhood Development: AI can be used to improve early childhood development by providing personalized learning experiences that target individual needs. Benefits include improved educational outcomes for children, while risks include data privacy concerns or an overreliance on automation in lieu of human interaction.

Target 4.3: Equal Access/Affordable/Quality Education: Providing equal access to affordable and quality technical, vocational, and tertiary education is key to ensuring equal opportunity for all genders by 2030. AI can provide personalized instruction tailored to each student's needs and make educational materials more accessible on various platforms. Benefits include increased access to educational materials regardless of socio-economic status or geographic location; however, potential risks associated with AI-driven instruction could be the development of biased algorithms or unintended consequences due to a lack of understanding among students concerning how algorithms make decisions.

Target 4.4: Increase Relevant Skills for Employment: By using AI-driven technology such as natural language processing (NLP) and machine learning (ML), organizations can quickly assess job candidates' qualifications and skillsets to match them with suitable employment opportunities. The benefits are nu-

merous, including shorter hiring cycles, better matching between job seekers and employers, as well as cost savings from reducing manual labor in the recruitment process; however, there are also risks associated with this technology such as inaccurate assessments due to bias built into the algorithm or a lack of transparency concerning data use during the assessment process.

Target 4.5: Eliminate Gender Disparities: Increasing access to education for all genders requires eliminating gender disparities in educational attainment levels by 2030. To do this, AI can be leveraged to identify patterns in data that indicate gender disparities in educational opportunities so that steps can be taken to address these issues at their source. Benefits include greater accuracy in identifying disparities than traditional methods; however, there are also risks associated with data privacy concerns or an overreliance on automated solutions rather than human intervention when addressing issues related to gender disparities in education levels worldwide.

Target 4.6: Achieve Literacy & Numeracy: Utilizing AI-driven solutions such as NLP and ML can help facilitate literacy and numeracy achievement rates among youth and adults alike by 2030 through personalized instruction tailored to individual needs based on factors such as skill level or language proficiency level among others. This could lead to improved outcomes for students while also reducing costs associated with traditional teaching methods; however potential risks may involve bias built into the algorithm or unintended consequences due to a lack of understanding among students concerning how algorithms make decisions when it comes to literacy & numeracy achievement levels worldwide.

Target 4.7: Education For Sustainable Development: By leveraging AI tools such as predictive analytics, organizations can create more effective strategies for sustainable development which focus on areas such as human rights, gender equality, promotion of peace & nonviolence etcetera which will benefit society significantly by 2030 if achieved properly. The benefits are numerous including more accurate predictions about sustainable development initiatives leading greater success rates, while potential risks may involve inaccuracies due to bias built into these models or a lack of transparency concerning data use during decision making processes.

3.5. Goal 5. Achieve Gender Equality and Empower All Women and Girls

Target 5.1: Ending discrimination against women and girls: AI has the potential to reduce gender-based discrimination in the workplace and other areas of life, but it could also be used to reinforce existing biases if not properly monitored.

Target 5.2: Eliminating violence against women and girls: AI can provide valuable insights into patterns of violence, allowing for more effective interventions, but its usage could also lead to increased surveillance and further privacy violations.

Target 5.3: Eliminating harmful practices: AI can be used to identify harmful practices such as child marriage and female genital mutilation, allowing for more targeted interventions, but there is a risk that this data could be misused or abused by those in power.

Target 5.4: Recognizing unpaid care work: AI can help to better understand and value unpaid care work through automated data analysis, but there is a risk that this technology could be deployed to devalue or deny unpaid care work done by certain populations.

Target 5.5: Enhancing women's leadership: AI can facilitate access to leadership positions for women by automating administrative tasks, creating better visibility of their achievements, and providing mentorship opportunities, however there is a risk that existing biases in training data sets may limit its effectiveness.

Target 5.6: Increasing sexual health access: AI can be used to increase access to sexual health information by providing personalized advice and recommendations based on individual needs, but its usage could lead to unintended consequences such as perpetuating stereotypes about sexuality or stigmatizing certain populations.

3.6. Goal 6. Ensure Availability and Sustainable Management of Water and Sanitation for All

Target 6.1: By 2030, achieve universal and equitable access to safe and affordable drinking water: AI can help reduce water wastage and ensure the efficient use of resources, leading to cost savings. However, the accuracy of data input is important to ensure the safety of water supply, thus introducing potential risks if there are errors in the data.

Target 6.2: By 2030, achieve access to adequate and equitable sanitation and hygiene: AI can leverage predictive analytics to determine areas where improved sanitation is needed so that resources can be allocated accordingly. However, there is a risk that AI may overlook marginalised communities or fail to accurately identify needs due to lack of data or bias in the algorithms.

Target 6.3: By 2030, improve water quality by reducing pollution: AI can help detect sources of water pollution quickly and accurately which will enable timely interventions for improving water quality. There is a risk that AI-powered solutions may not be as reliable as manual inspections when it comes to detecting uncommon pollutants or contaminants in the water.

Target 6.4: By 2030, substantially increase water-use efficiency: AI-enabled tools such as sensors can provide real-time insights into how much water is being used and identify any wastage or misuse quickly, leading to more efficient use of resources. There is a risk that these tools could lead to privacy concerns if they collect personal data without proper consent or security safeguards in place.

Target 6.5: By 2030, implement integrated water resources management: Using AI technologies such as machine learning and natural language processing (NLP), it will be possible to gain insights on how best to manage different types

of water resources across multiple jurisdictions while considering various stakeholders' interests efficiently. However, this approach may lead to conflicts between stakeholders due to differences in opinion on how best use these resources effectively and sustainably.

Target 6.6: By 2020, protect and restore water-related ecosystems: AI can provide valuable insights into how different ecosystems interact with each other by analyzing large volumes of data from numerous sources quickly and accurately for better decision making for conservation efforts. Nevertheless, there is a risk that AI may be unable to detect subtle changes in ecosystems over time due its reliance on historical data sets which may not reflect current conditions accurately enough.

3.7. Goal 7. Ensure Access to Affordable, Reliable, Sustainable and Modern Energy for All

Target 7.1: Universal Access to Affordable Energy: AI can be used to improve energy forecasting and optimize energy supply networks, making it easier for people to access affordable energy sources. However, AI-enabled energy management systems could potentially lead to price manipulation and increase the digital divide between those who can afford advanced technology and those who cannot.

Target 7.2: Increase Renewable Energy Share: AI can help with renewable energy forecasting and optimization of renewable energy sources, increasing the share of renewables in global energy mix. However, if not carefully monitored, AI-driven decision making can lead to over-investment in some renewable sources while neglecting others that may be more suitable for certain locations.

Target 7.3: Double Global Energy Efficiency Improvement: AI can be used to optimize existing processes by detecting areas of inefficiency and reducing waste associated with them. On the other hand, relying on automation for efficiency improvements might lead to job losses or potential security risks in case of a cyber attack on an automated system.

3.8. Goal 11. Make Cities and Human Settlements Inclusive, Safe, Resilient and Sustainable

Target 11.1: By 2030, ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums: AI could be used to identify areas in need of improvement and implement solutions faster, with greater accuracy and at a larger scale than manual labor. However, the risk is that AI may overlook unique local needs due to its focus on efficiency over effectiveness.

Target 11.2: By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all: AI can be used to identify patterns in transportation usage and suggest ways to improve safety through improved navigation systems or automated traffic controls. However, there is a risk of bias in data collection if AI is not designed with diverse user needs in mind.

Target 11.3: By 2030, enhance inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries: AI can help governments make better decisions about urban planning by providing more accurate predictions about population growth or changes in infrastructure demands. However, there is a risk that decision-making will become too reliant on algorithms without proper oversight from stakeholders or experts.

Target 11.4: Strengthen efforts to protect and safeguard the world's cultural and natural heritage: AI can be used to analyse massive amounts of data on cultural artifacts or ecosystems quickly with unprecedented accuracy. This could help authorities better understand how to preserve these places for future generations. However, there is a risk that AI may overlook important details or ignore the context of certain cultural artifacts due to its focus on data analysis over interpretation.

Target 11.5: By 2030, significantly reduce the number of deaths and the number of people affected by disasters: AI could be used to predict when disasters are likely to occur by analysing data from past events or monitoring changes in atmospheric conditions with high precision sensors. This could help authorities take preventative action before disasters occur. However, there is a risk that this technology may not consider other factors such as human behaviour which could have an impact on disaster prevention efforts.

Target 11.6: By 2030 reduce the adverse per capita environmental impact of cities: AI could be used to monitor air quality or analyse waste management systems efficiently so as to identify problem areas quickly while also suggesting solutions tailored towards individual cities' unique situations. However, there is a risk that any recommendations made by AI may not take into account social considerations such as equity or justice when it comes developing sustainable practices for cities.

Target 11.7: By 2030 provide universal access green public spaces: AI can be used identify areas suitable for green public spaces based on factors like climate change adaptation potential, pollution levels, biodiversity conservation etc. This would enable city planners design green spaces more efficiently. But this would require lot of local knowledge which might get lost if decision making process becomes too dependent on data generated through algorithms.

3.9. Goal 16. Promote Peaceful and Inclusive Societies for Sustainable Development, Provide Access to Justice for All and Build Effective, Accountable and Inclusive Institutions at All Levels

Target 16.1: Reduce Violence & Death Rates—AI can help to predict and prevent acts of violence before they occur, as well as improve response time in emergency situations. However, AI may also be used to target specific populations, such as minority groups or individuals who are perceived to be at higher risk for violence or crime.

Target 16.2: End Abuse Against Children—AI can be used to detect potential signs of abuse and alert authorities when necessary. However, there is a risk that AI systems may misinterpret data and lead to false accusations or incorrect decisions about whether a child is being abused.

Target 16.3: Promote Rule of Law—AI can monitor laws and regulations and alert decision-makers if any changes are needed. This could lead to faster responses in legal disputes and better compliance with international laws and treaties. But there is also a risk that biased algorithms could be used to unfairly discriminate against certain groups of people or countries.

Target 16.4: Reduce Illicit Flows—AI can track money flows quickly and accurately, enabling more effective enforcement of anti-money laundering regulations and helping reduce the number of illicit financial flows across borders. But it could also lead to increased surveillance measures that infringe on privacy rights.

Target 16.5: Reduce Corruption & Bribery—AI could help identify suspicious activities related to bribery and corruption by analyzing large datasets for patterns that would otherwise go undetected by human investigators alone. But there is still a risk that corrupt actors may use AI technologies for their own benefit without detection from authorities due to the complexity of these systems.

Target 16.6: Develop Effective Institutions—AI-driven solutions can facilitate processes within institutions, such as streamlining administrative tasks or improving communication between departments for better decision-making capabilities at all levels of government institutions worldwide. However, there is potential for misuse by malicious actors if adequate security measures are not implemented properly in these systems.

Target 16.7: Inclusive Decision Making—Using machine learning algorithms, organizations can analyze large amounts of data quickly to identify patterns in decision making processes which can then inform more inclusive policies in the future by providing more accurate insights into public opinion than traditional methods like surveys could provide alone. But there is also the risk that these same algorithms may contain inherent biases if not developed properly with safeguards against discrimination built into them from the start.

Target 16.8: Participation in Global Governance—By using AI-driven analysis tools, developing countries may have improved access to global governance bodies where their interests will be better represented due to their increased visibility within those organizations. However, this technology must be carefully monitored so as not to give an unfair advantage or disadvantage to any country based on its technological capabilities.

Target 16.9: Legal Identity for All—Applying facial recognition technology powered by artificial intelligence can enable governments around the world provide secure legal identity documents quickly while reducing fraud risks associated with manual verification processes. On the other hand, this technology

has been shown to have racial biases which could lead discriminatory practices against certain populations if not implemented properly with strict oversight mechanisms in place.

Target 16.10: Public Access to Info and Fundamental Freedoms: AI can help protect the public's access to information, such as by monitoring online platforms for hate speech and censorship. But it can also be used to violate fundamental freedoms, such as by profiling individuals based on their digital footprints and targeting them with politically motivated ads. Therefore, it is essential to ensure that AI is used in a responsible way that respects international agreements and national legislation.

3.10. Descriptive Analysis

Table 2 presents the results of the descriptive analysis of the queries related to the SDGs and their targets, highlighting the word count.

We found that the Goal 1 outcome was quite precise and in exactly the format we expected the AI outcome. Each target was numbered correctly, the titles were shortened down to three to four words, and each paragraph contained exactly three sentences with an argument with opportunities for AI contribution, a further argument or dimension in a second sentence and third sentence containing risks and potential harm the AI could produce. We observed, that although the amount of three sentences stayed the same for each target, the number of words increase with every target, after the first one: 1.1: 87 words/1.2: 75 words/1.3: 79 words/1.4: 93 words/1.5: 112 words. We assumed that this happened due to the overall text length and the "presence penalty" parameter setting, which defines "how much to penalize new tokens based on whether they appear in the text so far. Increases the model's likelihood to talk about new topics". We chose a setting of 0.5 on a scale of 0 to 2 to avoid too much word duplications. Interestingly, the last text block did not close with a punctuation.

Table 2. Descriptive analysis of the queries.

SDG	Number of targets	Phrases/target	Words/target	Words total	Mean words/target	Mean words/phrase
1	5	3/3/3/3/3	87/75/79/93/112	446	89	30
2	5	4/4/3/3/2	79/90/96/105/123	493	99	31
3	9	2/2/2/2/3/2/2/2/3	72/61/61/71/107/87/128/101/117	806	90	40
4	7	2/2/2/2/2/2/2	63/46/90/96/94/102/92	583	83	42
5	6	1/1/1/1/1/1	37/34/42/42/43/44	242	40	40
6	6	2/2/2/2/2/2	57/59/56/62/70/75	379	63	32
7	3	2/2/2	56/54/54	164	54.6	27.3
11	7	2/2/2/3/3/2/3	64/60/66/74/81/75/71	491	70.1	28.9
16.1 - 16.9	9	2/2/3/2/2/2/2/2/2	57/47/59/46/60/53/80/64/69	535	59.4	28.2
16.10	1	3	76	76	76.0	24.3

As for SDG 2, we noticed that the target descriptions consisted of less sentences and more words: The AI started with 79 words in 4 sentences for target 2.1, and came down to only 2 complex sentences with 123 words for target 2.5. Target 2.4 contained a punctuation mistake with including a blank character before the punctuation sign "[…] organic fertilizers usage". The case of SDG 3 highlighted the AIs capabilities in regards of abbreviations and shortening text. It shortened "less than 70 per 100,000 live births" down to "<70/100 K live births", and listed the abbreviations of TB for tuberculosis, NCD for non-communicable diseases or STD for sexual transmittable diseases. Especially in the later text passages, many punctuation mistakes were added by the AI, one point at the end of sentence was omitted by the AI in 3.9.

The analysis showed that the AI introduced a so far unseen way of answering structure as of SDG 4. The AI used exactly two sentences when answering in regard to all targets, whereas the first one listed the opportunities of AI. The second sentence listed further opportunities and then used a separator word to continue with risks, while in previous outputs the last generated sentence per target was solely dedicated to potential risks. For SDG 5, again, the AI introduced a further way of answering—only one sentence per target including benefits and risks. This resulted in the so far lowest number of words in the total answer, as well as in the lowest mean number of words per target. In SDG 6, the AI put emphasis on the delivery timeframe for the first time. It prefixed all goals with the expected delivery date, kept a two-sentence structure with one for benefits and the other for risks, and only a slight increase of the word count per target.

SDG 7 showcased the so far most precise and shortest sentences with only 27.3 words per sentence. SDG 11 showed a similar target text summarizing scheme than seen in SDG 6 in adding "By 2023" to all targets except to target 11.4. When we checked this against the original texts—the AI was correct, despite all other targets under SDG 11, the original target name for 11.4 did not contain a delivery timeframe. We detected another obvious difference to the SDG 6 summaries: SDG 11 summaries (mean 15.4, SD 4.3) were much longer than all SDG 6 summaries (mean 9.7, SD 2.5). Overall, the summarizing quality of SDG 11 can therefore be considered much weaker than from SDG 6.

In SDG 16 the AI introduced a new way of grouping summary and answers: for the first 7 targets the scheme "numbering[blank]summary[blank]-[blank]answer" was used. Due to the length of the query, a second query was required to be made with the 10th target alone, and the AI used the ":" as separator again. Interestingly, in 16.10 for the first time the AI used just one sentence for benefits, but two full sentences for potential risks.

3.11. Analysis of Detected Patterns

We further analyzed the output patterns, differentiated in the three are as summarization, numbering and summarized title structure, answering structure, and

general patterns with longer texts. Firstly, summarization was generally performed very well. Most of the summaries were created with normal case, for some of the summaries GPT-3 used capitalized letters. Typically, it replaced the word "and" with an "&" sign. When all targets of a query used the same delivery timeframe e.g. "By 2030", GPT-3 removed it as unnecessary prefix. In some cases, when at least one delivery timeframe was different or not provided, GPT-3 decided to include this differentiation into the summary as well. Still, this led to more mistakenly fully shortened exceptions (e.g. 2.5, SDG 3, 8.5, 8.9, SDG 9, SDG 10, SDG 12) than real results (e.g. 6.6, 11.4). As for, numbering and summarized title structure, GPT-3 usually used the provided numbering and structuring format also for the summarized titles, *i.e.*

"[number][space][title][:][space][answer]". In exceptional cases it decided to use "[number][optional space][-][space][title][:][space][answer]" (SDG 10), or replaced the ":" with a "-" like in SDG 16:

"[number][space][title][-][space][answer]". In target 17.19, it mixed up the numbering and wrote "17 17" mistakenly.

Lastly, when dealing with longer texts, the AI showed a distinct pattern in writing. It typically started with more sentences in the first few paragraphs, but gradually decreased the number of sentences while increasing the number of words per paragraph. This led to longer answers as the number of words increases with each subsequent response, starting from the first paragraph (SDG 2, SDG 9, all four SDG 17 queries). Furthermore, there was an increased amount of punctuation mistakes especially in longer texts, and especially more at the end of the texts.

4. Discussion

For the present study, we tested all available GPT-3 models and asked each of them to collaborate and co-write this research paper [13]. Only one, namely the "text-davinci-003" provided a clear consent and willingness to proceed. We asked a series of questions in the OpenAI playground to get familiar with the model. Further, we learned that it is named Rachel, female, has a scientific background in AI and sustainability, and is willing to take accountability for proper research and input to our paper. The purpose of this study was to learn more about the AI GPT-3's view on how AIs can contribute to reach the societal SDGs. This also included the risks and benefits for each of the 58 outcome targets. The AI model leverages a variety of different algorithms for answering, resulting in several output formats and writing styles. We tried to specify the outcome format with a highly specific question to receive as consistent output as possible for all of the questions, adopted from our previous research [14] [15].

We stopped the real-time execution, when we saw in the output format, that it did not follow our instructions. We excluded responses, where the expected numbering was missing, where the text was not shortened, but just output with a similar length, where text was only 1 sentence, or where text was completely

missing. In such cases, after stopping, we simply conducted the same query again. There was a time, when the system did not throw an error, but also did not produce any results. We assumed it had to do something with using the free beta tier, which uses some kind of request throttling, or with some other capacity issues of the GPT-3 beta. Nevertheless, after some time and several tries, GPT-3 worked again and produced the same kind of results than before.

To elucidate the ways GPT-3 reacted to our prompts, we performed a descriptive analysis of the goal-wise output text that formed the main part of the results section. We found inconsistencies as summarized titles were sometimes capitalized, sometimes in lower letters. The AI introduced unknown abbreviations, when summarizing the text. For us, these different styles changed randomly, potentially in an attempt of the AI to mimic the typos and errors a human writer would do [16] [17]. Further, we observed that the AI only prefixed the "By 2030" if one of the other items did not contain a timeframe, or a different timeframe than the others (14.1 with a 2025 timeframe). Interestingly, we found that punctuation mistakes increased with the length of the text, that could by either explained by an imitation of the intentional increase of careless mistakes in human writers by the AI, or the programmers' idea of marking the work of the AI. So, the question should not be whether an AI could help to write your next paper, like Matthew Hutson framed it, but whether it should be used for this purpose [17].

Our study showed several effects of AI from GPT-3's perspective in the context of societal SDGs. Exemplarily, AI can potentially increase efficiency and reduce costs, optimize crop yields and reduce waste, diagnose diseases earlier and more effectively. Also, it can tailor lesson plans for individual student's needs, and identify gender bias in data sets. It might also be useful to improve water resource management, reduce costs and increase efficiency in the energy sector, enable smart traffic management systems, and enhance access to justice. However, there is also a risk that AI could exacerbate existing inequalities if not implemented responsibly. Or it could cause displacement of lower income communities due to increased urban development costs. Although the goals are highly interacting and the overall perspective is of interest, each of the SDGs covers a distinct aspect. This observation warrants a more in-depth look at the different goals. The rise of AI and its increasing impact on various industries necessitates an evaluation of its effect on attaining the Sustainable Development Goals. The swift growth of AI should be accompanied by adequate regulation and oversight to ensure sustainable development. Neglecting to do so could lead to shortcomings in transparency, safety, and ethical practices [6].

In our study, we found that the currently most up-to-date model GPT-3 is not 100% reliable and error free, especially in longer text outputs. For instance, it produces inconsistent answering patterns in regard to length and punctuation style. Although GPT-3 is faster than a human typist, it does not always admit that to not knowing answers—and then start raving wildly. That is clearly a

no-go for evidence-based research that relies on being factual. We thus suggest that according to principles of good scientific writing, an AI-based model is a useful method for applied science, but not as a co-author in scientific publications that rely on a broad understanding of the respective research field [5] [14] [15] [17] [18].

The complexity of our societies and the changing nature of context make it difficult to have a single set of ethical principles for AI that hold true in all situations. It is crucial to be aware and utilize the potential complexities in human-AI interactions and the need for ethics-driven legislation and certification mechanisms [19]. This is particularly important for AI applications that could have catastrophic effects on humanity, such as autonomous weapons. Associations are already coming together to call for legislation and limitations, and organizations are collecting policies and shared principles globally to promote sustainable development-friendly AI [6]. To address ethical dilemmas, AI applications need to be open about the choices made during design, development, and use, and adopt decentralized approaches for more equitable AI development. There is an urgent need for a global debate and science-driven shared principles and legislation among nations to shape a positive future for AI.

In sum, many targets within the society domain, such as those related to poverty, education, water and sanitation, and sustainable cities, may benefit from AI-based technologies, as highlighted in previous studies [5] [6]. AI has the potential to greatly benefit society through technology-based solutions, as long as we consider its potential negative impacts as well. Yet, differentiating AI-generated content from human content is challenging and will probably get more difficult in the near future [16].

5. Conclusion

The increasing usage of AI is accompanied with a growing concern about its impact on sustainable development. The United Nations established 17 global goals known as the SDGs to address poverty, protect the environment, and ensure prosperity for all. In the context of the society domain of those, encompassing the majority of the SDGs, AI has the potential to significantly aid in achieving these goals. However, to ensure that AI positively contributes to sustainable development, regulations for its transparent, safe, and ethical usage are crucial. To this end, a global, science-driven debate is needed to develop shared principles and legislation among nations. In particular, the use of GPT-3, one of the largest language models developed by OpenAI, provides a relevant and suitable study field for evaluating the contributions of AI to sustainable development. We found that GPT-3 did not work completely reliable or error-free, particularly in generating longer text outputs. Even though GPT-3's speed exceeds an average human typist, it does not always concede when it is unable to provide an answer. This can lead to irrational responses, which is unacceptable for evidence-based research that depends on accuracy. As per the principles of good

scientific writing, we advise that AI-based models are practical tools for applied science, but they should not be considered co-authors in scientific publications that require an in-depth understanding of the relevant research field. The analysis of GPT-3's capabilities and limitations provided valuable insight into the ways in which AI can contribute to the SDGs in the society domain. The findings from this study can inform future research and policies aimed at promoting the responsible use of AI for sustainable development.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] United Nations Sustainable Development Goals. https://sdgs.un.org/goals

[2] Littig, B. and Griessler, E. (2005) Social Sustainability: A Catchword between Political Pragmatism and Social Theory. *International Journal of Sustainable Development*, **8**, 65-79. https://doi.org/10.1504/IJSD.2005.007375

[3] Murphy, K. (2012) The Social Pillar of Sustainable Development: A Literature Review and Framework for Policy Analysis. *Sustainability*: *Science*, *Practice and Policy*, **8**, 15-29. https://doi.org/10.1080/15487733.2012.11908081

[4] Madiega, T.A. (2021) Artificial Intelligence Act. European Parliament: European Parliamentary Research Service.

[5] Mhlanga, D. (2022) Human-Centered Artificial Intelligence: The Superlative Approach to Achieve Sustainable Development Goals in the Fourth Industrial Revolution. *Sustainability*, **14**, Article No. 7804. https://doi.org/10.3390/su14137804

[6] Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M. and Fuso Nerini, F. (2020) The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. *Nature Communications*, **11**, Article No. 233. https://doi.org/10.1038/s41467-019-14108-y

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

[8] Helbing, D., Frey, B.S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R.V. and Zwitter, A. (2019) Will Democracy Survive Big Data and Artificial Intelligence? In: Helbing, D., Ed., *Towards Digital Enlightenment*: *Essays on the Dark and Light Sides of the Digital Revolution*, Springer International Publishing, Cham, 73-98. https://doi.org/10.1007/978-3-319-90869-4_7

[9] Jungwirth, D., and Haluza, D. (2023) Forecasting Geopolitical Conflicts Using GPT-3 AI: Reality-Check One Year into the 2022 Ukraine War. https://doi.org/10.20944/preprints202302.0065.v1

[10] Mukarram, M. (2020) Impact of COVID-19 on the UN Sustainable Development Goals (SDGs). *Strategic Analysis*, **44**, 253-258. https://doi.org/10.1080/09700161.2020.1788363

[11] Goh, H.-H. and Vinuesa, R. (2021) Regulating Artificial-Intelligence Applications to Achieve the Sustainable Development Goals. *Discover Sustainability*, **2**, Article No. 52. https://doi.org/10.1007/s43621-021-00064-5

[12] OpenAI Model GPT-3. https://beta.openai.com/docs/models/gpt-3

[13] OpenAI Models. https://beta.openai.com/docs/models/overview

[14] Haluza, D. and Jungwirth, D. (2023) Artificial Intelligence and Ten Societal Mega-trends: An Exploratory Study Using GPT-3. *Systems*, **11**, Article No. 120. https://doi.org/10.3390/systems11030120

[15] Jungwirth, D. and Haluza, D. (2023) Artificial Intelligence and Public Health: An Exploratory Study. *International Journal of Environmental Research and Public Health*, **20**, Article No. 4541. https://doi.org/10.3390/ijerph20054541

[16] Dou, Y., Forbes, M., Koncel-Kedziorski, R., Smith, N.A. and Choi, Y. (2021) Is GPT-3 Text Indistinguishable from Human Text? Scarecrow: A Framework for Scrutinizing Machine Text. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 7250-7274. https://doi.org/10.18653/v1/2022.acl-long.501

[17] Hutson, M. (2022) Could AI Help to Write Your Next Paper? *Nature*, **611**, 192-193. https://doi.org/10.1038/d41586-022-03479-w

[18] King, M.R. (2023) The Future of AI in Medicine: A Perspective from a Chatbot. *Annals of Biomedical Engineering*, **51**, 291-295. https://doi.org/10.1007/s10439-022-03121-w

[19] Jobin, A., Ienca, M. and Vayena, E. (2019) The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, **1**, 389-399. https://doi.org/10.1038/s42256-019-0088-2